A Stroll Through Physics
BOOK TWO

David L. Peterson

"We shall never cease to stand like curious children before the Great Mystery into which we were born."

"Out yonder there was this huge world, which exists independently of us human beings and which stands before us like a great, eternal riddle, at least partially accessible to our inspection and thinking. The contemplation of this world beckoned like a liberation…"

"That which is impenetrable to us really exists, manifesting itself as the highest wisdom and most radiant beauty, which our dull facilities can comprehend only in the most primitive forms. Veneration for this force beyond anything that we comprehend is my religion."

Albert Einstein

Dave Peterson, 4350 Drew Circle, Boulder, Colorado, 80305
Email: davepeterson137@gmail.com
Date: December, 2019 {With special thanks to Mike Jones, *Dr. R.Michael Jones, Boulder, Co.*}

Contents, Book 2:

[December, 2019]

David L. Peterson , A Stroll Through Physics II,

1. [Content Listing of **Book 1**, 2014 at www.Sackett.net/DP_Stroll.pdf 303 pages].
{There is also a ***Book 0,*** *General Relativity, Dave Peterson 1979 }* *240 pages.*

2. "**Introduction**" December, 2019. (23 pages)
3. "Learning Quantum Mechanics and Relativity," May 2015, (26 pages)
www.sackett.net/LearningQuantumMechanics.pdf
4. "Physics Lives in Form Heaven," April, 2015, Posted at **FQXI** Foundational Questions Institute Essay Contest [http://**fqxi.org**/community/forum/category/2351 (9 pages)
5. "What Fundamental Should Mean," **FQXi** Essay 12/14/17
https://fqxi.org/community/forum/category/31426 (10 pages)
6. "Photons and Light" 3/6/19 (14 pages)
7. "Explaining S Orbitals and Bonding," Aug. 2018 (13 pages)
8. "Reality of Schrodinger Ψ-Waves, 9/14/18 (5 pages)
9. "An Electron is Waves of What?" 11/1/19 (11 pages)
10. "Concrete Hidden Variables," December, 2016 (16 pages)
11. "deBroglie-Bohm Interpretation of Quantum Mechanics 5/30/10 (9 pages)
12. "Appearances of Retrocausality in Entanglement Experiments," (9 pages)
13. "The Density Matrix," May, 2014 (8 pages)
14. "Quantum Measurement," 7/7/15 (6 pages)

15. "Nobel Prize for Topology in Exotic Materials," December, 2016 (17 pages)
http://sackett.net/TopologyInExoticMaterials.pdf
16. "Spinors" 6/10/18 (30 pages)
& Properties of Bi-Quaternions 1976 (4 pages)

17. "Lie Group Representations," June, 2015 (8 pages)
18. "Geometry in Modern Physics," July, 2016, (32 pages)
19. "Covariant Derivative Issues," September, 2017, (19 pages)
20. "The Lie Derivative" 6/22/19 (19 pages)
21. "Long Spin Disk Lube Migration, I & II," *Favorite Publication,* (11 pages)
DP, IEEE Transactions on Magnetics, Vol 28. No. 4., July, 1992
+ Part II, Vol 32, No. 5., Sept 1996,

22. "Tape Magnetism Waveforms" 5/29/01+ (11 pages)
23. "Brief Summary of Collapse to a Schwarzschild Black Hole,"
October, 2017 (9 pages)

24. "Boulder Cosmology Group Comments." Nov. 2019
(Dave Peterson, 2/15/16 to present). (27 pages)
25. "Highlights of the Months," Physics, 2014-2017 (43 pages)

Dave Peterson, 4350 Drew Circle, Boulder, Colorado, 80305

[Σ =391 pages]

Email: davepeterson137@gmail.com

CONTENTS of Book One:

May 9, 2014, dp

“A Stroll Through Physics.pdf’’

NEW REPORTS > 2010 [LaTeX]:

GENERAL RELATIVITY/COSMOLOGY:

``Graphical Representation of Radial Coordinates in Cosmology, '' DP, 10 January, 2011, 6 pages. (Graphical Representation.txt, from 1979.).

``Decomposition of the Perihelion Shift of General Relativity,” DP, 31 Jan 2011, 13 pages. (PN_Decomp.txt)

``The Radius of the Universe,” 27 February, 2011, DP, 9 pages. (Radius Universe.txt)

``Cosmological Distances,'' 21 April, 2011, DP, 17 pages, (10/5/11, Cosmology_1.tex).

``Comments on Inflation,'' DP, 11/10/2011, 9 pgs.

``Hawking Radiation,'' 9/6/13. 12 pgs,

``Recent Results in Astrophysics,'', 8/28/13.(CMB, AMS, light, galaxy, WDM, Fermi Bubbles). 10 pgs

PARTICLES/QUANTUM:

``The Last Decade in Experimental Particle Physics'' updated to 2/9/12, 11 pgs.

``Rotations of Base States,'' DP, 11/20 – 12/15/11 5 pgs.

``The Fine Structure Constant,'' DP 11/20/2011, 6 pgs.

``Electron Spin and SU(2),'' 9/26/12, 15 pgs. (Spin.txt).

``Explaining S Orbitals and bonding,'' DP, 6/24/2012.pdf to 3/6/13. 11 pgs.

``Test of Quantum Entanglement—Aspect Experiment,'' 21 June, 2011, 9 pages (word doc), to 6/26/11.

``Beneath Quantum Mechanics,'' DP 15 March, 2011, 42 pages. to 7/12. (Underlying QM.txt)

``WaveFunction Sub-Quantal Information,'' 11/20/13-3/24/14, 32 pages.

``Gauge Theory for Electromagnetism,” DP, November 6, 2011, - 12/22/11 . 26 pages,

``Five Dimensional View of Electricity and Magnetism,'' 9/15/05, 2010, 8 pgs.

``Circle Models in Modern Physics,'' 3/6/2012, 10 pgs.

`` Special Topics'', water waves, matter wave index 8/21/12.

OLDER WORK: (Appendix/Optional   ~ 50 pages).

``Simple Mnemonic Device for Nuclear Shell Filling,’’   27 Sept. 1968 (for Am.J.P).

``An Interesting Function’’ (to teaching journal). 1983, $\rho = \epsilon E1 \cdot E2$.

``General Relativity for Pedestrians,'' First order approximation, 1974, DP.
<u>THE GENERAL THEORY OF RELATIVITY</u>, sample pages and listed contents from a proto-manuscript of 1979. [*240 pages*]. Dave Peterson.

``Consistency of a Mechanism for Mach's Principle,'' 1969.
Sample: ``An Analogy between the Linearized theory of General Relativity and Electromagnetism,'' August 1976.
``Einstein's Formula for Mach's Principle,'' 1977

<u>Career Highlights Resume of Engineering Experience.</u>
List of Company Publications *and Citations* (Storage Technology Corporation, STK).

<u>Sample First Pages of Key STK Publications [D.L. Peterson]</u>:

`Long Spin Disk Lube Migration,' 1992; And `Part II', 1996; IEEE.
`Disk Surface Interferometry,'
`Sputtering Through Offset Mask for Disk Acceleration Standards,' 1993;
`Laser Scribing on Magnetic Disc,' 1981, Computer Design.
`Error Theory for Laser Disk Standards,' 1984, IEEE.
`Controlling Magnetic Contamination in the Disk Drive Industry,' 1993, `
Sample US Patent #4, Tape Backhitch Time. 2006
` Power Laws in Large Shop DASD I/O Activity.' 95
`Fractal Patterns in DASD I/O Traffic,' [voted Best Paper, CMG96]
`Data Center I/O Patterns and Power Laws,'' CMG'96
`DASD Subsystem Cache Statistics,' 1997.
`New Perspectives in DASD Subsystem Cache Performance,' 97

`Power Laws in DASD I/O Activity and Cache,' 97

<u>Not Shown: *STK Symposium*</u>:
`Tape Track Mis-registration Modeling (TMR),' '99
`Buffer Design for Future Tape Drives,' '01
`Modeling for Operating Points of Proposed Tape Drives,' '01
`Tape Track Distortion,' 2000
`Modeling of Tape Errors and ECC Effectiveness.' '99
`Complexity in Computer Shops.'

Syllabus: CU Division of Continuing Education, Independent Study, Boulder.
General Physics Phys. 201-4, Phys. 202-4 © 1972-1977.
Physics For the Life Sciences, 201-L-4, David L. Peterson, © 1975.

``Elastic Scattering of 28 MeV Protons from Deuterium,'' NSF Summer Research Project, David L. Peterson, 57 pgs, 1965.

Introduction (Book Two).

Dave Peterson, 7/26/2019 -- 12/10/19.

My View of the Physical World

Overview:

I've been studying physics for the last sixty years with a goal of uncovering and comprehending its most basic fundamental laws -- the reductionist principles by which the universe functions from the inside out. This means focusing on the foundations underlying the wonderful facts of physics and stressing the "common denominators" of physical actions along with a pursuit of basic unanswered questions. Some of my most fundamental questions about physics have not yet been answered. But, I wish to declare my final guesses about such topics as: "natural mathematics," the future of improved interpretations of quantum mechanics, and the properties of the "Vacuum." In modern physics, the vacuum of space-time is no longer an empty nothingness. It is rather the most important essence that I refer to as "Plato's Form-Heaven" – the bearer of all the invariant physical constants, laws of Nature, and fundamental templates for fields and quanta.

Physics is primarily expressed in the language of mathematics. But we also want intuitive explanations of that math accompanied with the human history of physics so we may all share in the key "Aha ! " moments that changed our core beliefs about the world. Explanations depend on stating concepts but also on finding appropriate words with updated definitions for modern understandings. The more classical terms used to express science often have an unhealthy prolonged inertia that gets in the way of these new understandings. Mathematics has precise definitions, but definitions in physics evolve with new knowledge: what really is mass, energy, gravity, a "particle," a photon, an electron, quanta, measurements, fields, the Vacuum,… and most difficult of all, "reality" itself ?).

<u>This Book:</u>

I have now written three physics "books" – mainly just for myself and a few friends. The process and discipline of researching and writing makes key concepts more tangible, concrete and memorable to me and I hope more presentable to others. The essays document my understandings so that re-reading facilitates re-learning.

My self-study of general relativity from 1968 to 1980 was gathered together into what I call **"<u>Book Zero</u>."** Rather than expounding on the complexity of the subject, my major goal was to try to simplify the concepts for a more intuitive understanding. One could say for example that the bending of starlight, the time delay of radar, and gravitational red shift are simply consequences of special relativity combined with the principle of equivalence (SR+PE, [Schiff]) not needing the power of general relativity. And, the historically decisive "perihelion precession of Mercury" comes close to also being intuitive. In the realm of cosmology, our "flat" universe cosmic expansion is in part like a simple Newtonian model for expanding dust [Liddle]. But, when we get to concepts such as rotating black holes and gravitational radiation, then we do require the full theory.

The essays from about 2009 to 2014 that are presented in **Book One** were broader – still beginning with general relativity but moving more into modern physics: cosmology, Kaluza-Klein theory, understanding up-to-date particle physics and astrophysics, efforts probing the foundations of quantum mechanics, electrodynamics, nuclear shell filling, and a partial overview of my previous work-studies in applied magnetics.

An incidental note here is that introductory magnetics raises the curious question, "What is an intuitive understanding of the Lorentz force?" [$F = qv\times \mathbf{B}$]: a particle velocity one way crossed with a B field another way gives a force perpendicular to both – doesn't that seem strange? What other force field does that: the Coriolis force, $F = 2m\, v\times \boldsymbol{\omega}$. This (so-called fictitious force) is due to our being in the "wrong" frame of reference – a rotating frame. The $2m\omega$ part is $\sim 2m(v_\phi /r)$, and since B=curl A, $qB \sim q(2A_\phi/r)$, where A is the vector-potential, and q is the electric charge on a particle. The term "qA" acts as an "electromagnetic momentum" competitive with p=mv. So a region with magnetic field B acts like an unseen rotating frame of reference dragged along by A but only experienced by particles possessing charge. An electron sees this along with its usual inertia due to its mass (an "e" part plus an "m" part).

The new **Book Two** discusses mathematics used in modern physics such as that found in The Geometry of Physics [for example, the book by Frankel]. It begins with my essay on "Learning quantum mechanics and relativity" and my two essays for the "Foundational Questions Institute" [FQXi.org]. In order to dig under the present mathematics of physics, to find its commonality and simplifications and mappings to "reality," it is necessary to first have familiarity with some of the higher math. That includes topics such as differential geometry, differential forms, covariant differentiation, connections, curvature, topology, Lie groups, and Yang-Mills fields. My essays talk lightly about the appropriate math and physics used by Nature for quantum mechanics, entanglements, quantum optics, and a little more on relativity and cosmology.

A major interest is the possibility that hyper-complex numbers might best represent the mapping to sub-quantum reality in the arenas of quantum-field theory and particle physics. An intuitive reason for that is that the quantum world is generically like "the square root of reality" in the sense of ψ versus $|\psi|^2$. A special example is Dirac effectively taking the "square root" of the Klein-Gordon equation to get his equation for the electron. And Pauli matrices for electron spin are essentially complex quaternions which have three imaginary bases, i, j, and k. Blends of complex and real bases are covered by the term "Clifford Algebras."

My Purpose:

Apart from persistence and a deep appreciation of Nature, I am essentially a generalist and a perpetual and typically average graduate physics student. I care very much about elevating and improving the teaching of physics. Physics is a hard subject; but, with a change in style, it could be presented with more simple intuitive clarity and inspirational motivation. Should it remain as dry and difficult and impersonal as usually given in texts and articles? Many of us wish to be presented with "what is really going on" in the physical world, how does Nature really work, what are the bottom lines? I'm not aware that anyone can answer these questions, but we should try.

I have noticed that physics is addressing some of these concerns. For example, the first pages of many physics journals are more readable than they used to be. And the Physical Review guidelines now say, “Direct the manuscript text at a general readership, so as to make it understandable to a broad spectrum of researchers” avoiding jargon and the excessive use of acronyms. Page costs and length restrictions are no longer major constraints for web publications such as ArXiv.org, and that allows for much more flexibility in writing. There are a lot of popular books on modern physics, but almost all are at a low level that isn’t real enough to satisfy. More are needed at intermediate levels between too easy with inappropriate analogies and too hard without interpretation. Fortunately, we do have Physics Today and Physics World magazines.

One of the best experiments in writing style was The Feynman Lecture Series – a wonderful work with motivations and colloquial clarity. Feynman was unusually “real” and honest in stating that no one knows what is going on beneath quantum mechanics and that one shouldn’t try too hard – he didn’t forbid it but warned that thinking too much about it could be detrimental to career and mental health. However, his own early lack of understanding resulted in his creation of a new formulation of quantum mechanics referred to “sum of histories” or “path integral” – something that made sense to him where Copenhagen quantum mechanics didn’t.

I think that my paper on “Learning Quantum Mechanics and Relativity” is an example that shows how teaching could be improved. There are simple intuitive introductions to the dryly-stated logical and abstract postulates of quantum mechanics. There are simple explanations of what a “Lagrangian” really means. In relativity, a metric states the distances between two events as “deviations from the speed of light” (a light-like separation has zero proper time, $\Delta\tau = 0$) – a very different scenario from Euclidean geometry. That is a given, but why? And, in quantum mechanics, before presenting the complex math of the hydrogen atom, first do a simple exercise of trying out $\psi = Ae^{-br}$ in the Schrödinger equation. This yields the 1S orbital – the important orbital that Bohr missed.

There are many other basic simple and very inspiring explanations that I wish every student of physics would read. These heuristic examples include: elementary essentials in studying general relativity [Schiff], a scenario “On the Origin of Inertia” [Sciama], the transactional interpretation model of quantum mechanics [Cramer], and cosmology in terms of the Newtonian “expanding dust” model {e.g., [Pettini], [Liddle] }.

I wish to ease the learning pathways to avoid some of the unnecessary frustrations of physics students and the great frustration not being able to see or even imagine “what might be real” in quantum mechanics. And I always wish to dig deeper into the workings of Nature. Almost every topic I’ve studied in my essays has involved assembling heuristics found in a great many sources – it is rare that just one or two suffices. Finding answers to common questions about reality is usually an unanticipated struggle. I think this is a strange deficiency in our publications and teaching methods – why can’t we do better?

We are advanced biological creatures, a result of more than a billion years of evolution. As such, we have developed natural built-in drives and purposes: finding food, seeking safety and shelter, pair bonding, creating a family, belonging to a tribe or society and conforming to its beliefs, finding certainty, striving to achieving social status or power. Most people in most societies are able to find a path that gives their lives meaning. But now, for the first time in the history of people and civilization, science offers us the very new and different purposeful path of striving to see into the broad reality of our universe.

We are borne into a "Great Mystery" – an immensely huge, deep, intricate, intelligent and very strange physical world. Science has advanced enough so that we can now realistically reason about this factual world and begin to collectively make sense of its "magic." We can see the intellectual architectures of the cosmos and the micro-world that enable us to deeply appreciate a universal physical reality instead of just our usual, somewhat narrow, largely artificial and sometimes unpleasant every-day world. Pursuing this new goal is a singular deviation from previous life conventions and may provide a sense of freedom from otherwise feeling "programmed" by our genes and by our cultures. One can appreciate the latest scenes and mysteries at the edges of the new frontiers, and they reveal an incredible and "preposterous universe."

So, what my evolving knowledge does for me is provide a strong ongoing sense of amazement and mystery – a somewhat "spiritual" value that could be called "Deep Nature Appreciation." I imagine that is similar to what Einstein calls his "cosmic religion."

Along with learning physics is the major attraction and personal satisfaction from learning the "Big Story" of Physics – the set of all the individual biographies and discoveries of the key players – the great eye-opening moments of history that changed our view of reality. Facts should be coupled with physics also being considered as a Human pursuit. Unfortunately, there is a common principle of presenting physics as impersonal, logical, and concise so that most textbooks skimp or even ignore the history of science. But this history facilitates the understanding of the progress of science for us and is one of the meaningful joys of learning. Physics could and maybe should be taught more often through its history. It is a longer approach, but it sustains motivations. It is through history that we can re-live the great break-through experiences where we had lived in one world but then began to live in another broader and more enlightened world.

A classic example is the famous transitional figure of the 1600's, Isaac Newton, who has been called …

" the last of the magicians, the last of the Babylonians and Sumerians, the last great mind which looked out on the visible and intellectual world with the same eyes as those who began to build our intellectual inheritance rather less than 10,000 years ago [John Maynard Keynes]."

After his great work in the Principia, science entered into the paradigm of a Newtonian mechanical clockwork universe and stayed with it for nearly two centuries.

A partial list of the key transformations of our world-view in the history of physics could include:

The awareness of the extreme rapidity of light (Ole Romer, 1676), laws of mechanics (Isaac Newton, 1687), "light as a wave" (Thomas Young, 1814), electromagnetic field induction (Michael Faraday, 1831), conservation of energy (Hermann von Helmholtz, 1847), ineffectiveness of the "aether" concept (Michaelson,1887), little particles of charge (J. J. Thomson, 1897), the introduction of Planck's constant of action (Max Planck, 1900), special relativity (Albert Einstein, 1905), the Bohr atom model (Niels Bohr, 1913), gravitation is geometry (Einstein, 1915), discovery of the neutron (James Chadwick, 1932), matter waves and the Schrödinger equation (Erwin Schrödinger, 1926), antimatter (P.A.M. Dirac, 1928, 1932), The "Big Bang" universe (Georges Lemaitre, 1927), Hubble expansion (Edwin Hubble, 1929), dark matter (Jan Oort, 1932, Fritz Zwicky, 1933, Vera Rubin, 1970), the muon (Carl Anderson, 1937 – *"who ordered that! " still applies*), the extreme luminosity and redshift of quasar 3C273 (Maarten Schmidt, 1963), cosmic microwave background radiation ("CMB" 1964), pulsars (Jocelyn Bell, Antony Hewish, 1967), charm quarks (Richter and Ting, 1974, *the "November Revolution" of physics*), weak W and Z bosons (Carlo Rubbia, 1983), the accelerating universe (Perlmutter, Schmidt, Riess, 1998), and gravitational wave detection ("LIGO," 2015).

Problems needing Answers:

Physicists generally acknowledge a somewhat standard list of the leading important outstanding problems that need to be solved -- It includes:

What is Dark Energy? Why is the cosmological constant so tiny when vacuum zero point energy should be huge? What is Dark Matter? Why and how is there more matter than antimatter in the universe? What is the correct model for cosmic inflation? Why are there three generations of particles, and what underlies the strange values of masses of the elementary particles? Is there really a long-range barren desert below the list of currently known particles and energies? How does the neutrino get its tiny mass? Explain the "arrow of time." Prove QCD color confinement. Find a quantum theory of gravity. One might also add, "Why is there something rather than Nothing?"

But the list that bothers me most is much simpler, commonplace, and relatively unstressed by others:

1. Single photons can travel through complex optical glassware as if they were large amplitude classical electromagnetic waves. This is just amazing to me – but it is rare that it is mentioned at all in books or articles and just seems to be taken for granted !
2. Pauli exclusion is fundamental and consistent-with but not directly derivable-from quantum field theory (QFT). It holds up mountains; but there is no clearly understood reason for the spin-statistics relation. That is future physics.
3. The important and basic "Born Rule**"** (~ $\psi^*\psi$) is postulated but not fundamentally understood. The "psi" of QM or QFT lives in what I call "the square root of reality," and the concept and existence of a "probability amplitude" seems strange but is now taken for granted – it makes the world work.
4. Large objects like buckyballs self interfere in double slit interference despite the fact that they are much smaller than the separation between slits. Their aggregate mass $M=\Sigma m_i$ and its momentum become de-Broglie waves traveling through both slits. What is the mechanism for forming $\lambda = h/Mv$? Is it really a Lorentz transformation of a huge composite vibration: $\omega = Mc^2/h$?
5. If protons and neutrons in the nuclei were pictured as little balls of experimentally confirmed diameters (a la Linus Pauling models), they would be packed so closely together that no ball could interpenetrate them. Yet n's and p's (and α's) speed around fairly freely inside the dense nucleus and can organize themselves in the angular momentum "shell model" as if they were all "really" "waves" rather than particles. Explain this paradox (is it non-interaction due to different sets of quantum numbers)?

6. Most of the photons in our expanding universe will never encounter absorbers ("photons that never end") – but quantum optics almost defines photons as

existing when and where they are detected! The photons in the universe contribute to total energy – so do they exist as individual entities while they travel as rays? It is also true that neutrinos do not interact over cosmological distances in the later universe ("neutrinos that never end"). About 9% of cosmic rays are alpha-particles (α's that never end – are they rays?).

7. Non-locality and entanglement superpositions are a way to enforce conservation laws given a world of possibilities. In QM, "distance doesn't matter" and negligible amplitudes still function. Does this involve sub-quantum communication back-and-forth in time? (à la John Cramer?).
8. How can the realm of energy cover such an Incredible Range: Neutrino energy and photon energy has been measured up to 10^{11} and 10^{12} eV! It is believed that we might approach the "Planck energy." But photon energy can also go down below micro-electron-volts, μeV.

9. Does the spacetime Vacuum automatically take derivatives: for example, does Nature promote the vector potential, A, into energy bearing E and B fields (or E^2 and B^2) by effectively performing differentiations? $B = \nabla \times A$ and $E = -\partial A/\partial t$ (or for differential forms, F = **dA**). And then, for wave-functions, we add the operators $p = -i\hbar\nabla$ and $\mathcal{E} = +i\hbar\,\partial/\partial t$. There are other applications where 2nd derivatives are required such as the d'Alembertian operator $\Box$.
10. Is there a barrier to understanding Nature beyond which we will not and can not penetrate? – a permanently hidden functional "Core" forever beyond the ability of humans and their science (and our super-computers).
11. What are the key concepts that ultimately enable the emergence of "life?" Consider a newly formed planet Earth created from a dense primordial soup of atoms covering much of the periodic table. Then flash forward billions of years to the present age of humans: we exist and walk about and we encounter things like big trees with massive trunks. What are the general principles that enable that sort of life to architect itself and rise above the surface of the earth?

During the 20th century, progress in science and technology advanced exponentially and drastically changed our world and our worldview. Many even believe that the pace of change itself is advancing exponentially with a resulting hyper-exponential "law of accelerating returns." We essentially live in a totally different world from that of the previous century. We have few generalists and now mainly specialists who themselves often cannot keep up with their specialty. In our personal world, apart from the requirements of our employment, we can manage the great expanse of scientific knowledge by focusing on the smaller set of essential foundations – the facts from which the other facts can be derived. What are the basics? What are the foundational laws? What makes reality work? And, just how deep can we go? And are there some concepts beyond the ability of humans to understand or even probe.

Since about the year 2010, we have learned that the geometry of the cosmos is "flat" with nearly zero spatial curvature. Among other things, this means that cosmological theory unexpectedly became accessible even to striving laymen due to a minimal need for general relativity. We are now able to "cookbook" the evolution of the universe from the first microsecond to the present age nearly 13.8 billion years after the Big Bang. To a significant approximation, we now have an accepted "Standard" or "Concordance Model" of the universe. It is presently called flat "Lambda-CDM" and is composed primarily by a big dose of something called "dark energy" vacuum pressure along with a lot of invisible "cold dark matter" and only a small portion of our familiar ordinary matter. We don't yet know what dark energy is, but it presently seems to be Einstein's "cosmological constant" from 1917-- a λ or Λ "anti-gravity" term in the Einstein equations now causing accelerated expansion of the universe. And we don't yet know what non-interacting dark matter is, but we are actively striving to figure that out soon. We've made very impressive progress but obviously need to know more -- and maybe that will always be true.

In the realm of the microworld, we now have what we consider to be a complete list of the 62 elementary particles of physics along with the "Standard Model" of their field theories. And we recently verified the existence of a scalar particle called the "Higgs" as a quanta of an all-pervading "Higgs field" that causes the masses of most of the elementary particles. Despite this, there are indications that we will eventually need to go beyond the standard model ("BSM") to deal with such problems as neutrinos having tiny masses that had been predicted to be zero.

We have incredibly successful mathematical theories of quantum mechanics and quantum field theories that consistently pass all experimental tests. What we are presently missing is how to interpret them in ways we might comprehend or at least accept as humans -- and what lies <u>beneath</u> these theories that enables them to function as they do. Albert Einstein once said that the supreme task of physics is to arrive at those universal elementary laws from which a world-picture can be built up by pure deduction. The standard or "Copenhagen" paradigm of quantum mechanics only began to give way after the published works of David Bohm and John Bell. Bell said that the purpose of a theory is <u>to understand the physical world and</u> that "to restrict quantum mechanics to be exclusively about piddling laboratory operations is to betray the great enterprise." [Bell,1990]. And Bohm's alternative interpretation of quantum mechanics was realistic.

When I was twelve I read a basic book by Einstein that led me to want to devote my life to learning, understanding and appreciating how the universe works. That led me to declare a major in physics in 1960. Now, after sixty years of studying a large cross section of physics, how much "wisdom" do I have? Am I satisfied with how deeply I can see into the mysteries of Nature? Well, … … … …

I have learned a lot of fascinating stuff. Physics can see very far and certainly much deeper than we have ever possibly imagined. But, from our present perspective, there is certainly more that needs to be known, and Nature's fundamental reality sometimes appears to be unexpectedly complex and strange and seemingly opaque. The complexity of new developments grows faster than the ability of individual comprehension. So, I do not have all the answers I desired.

Should we be able to see deeply into the Spinozan "Substance" of Nature? The amazing progress of physics has shown that we actually can penetrate the ultimate to a remarkable degree in accordance with Einstein's statement that "god is subtle but not malicious." The main reason for our success so far is that the mathematics we have applied indeed seems to be an appropriate and powerful tool. And a lot of intentionally "pure mathematics" has unexpectedly found real application in physics. So a big question is "can our relevant mathematics be truly <u>isomorphic</u> to deep reality and finally go all the way?" Nature functions in a way that agrees with mathematical descriptions, but might there be an optimal one-to-one mapping to the way Nature actually operates? When a collective "intelligence" shines its light on Nature, it also discovers pre-existing mathematics. That is one of the reasons that mathematicians tend to be Platonists, many believe that even their pure mathematics is discovered rather than invented.

A further thought is, "might ultimate real mathematics be so deep and so complex that it is beyond human comprehension." We have had glimpses of this such as the following: "It may be that the fundamental operations of the standard model of particle physics are isomorphic to the hypercomplex octonians" – a proposal that has been seriously offered from time to time {e.g., [Atiyah],[Jackson], [Furey] -- *(octonians are a hyper-complex number system that uses* <u>*seven*</u> *imaginary numbers – seven different square-roots of minus one, -1 }.* Even though most of us couldn't handle this complexity, perhaps our super-computers or artificial intelligences could. An example of this is that Quantum ChromoDynamics (QCD) can be stated in short equations. But actually using it requires very long super-computer calculations on a finely spaced 4D spacetime lattice. If we have to have this level of high-tech assistance, can we still say that we understand?

And then, a further confounding thought is that although the mathematics of non-relativistic quantum mechanics ["QM"] is already well known and not too difficult, we still haven't been able to interpret what it means and what lies beneath it. So, again, is there a limit to human ability to comprehend the universe? I read a book by David Bohm on his view of quantum mechanics before I ever had a class on the standard theory. I always knew that I was hearing some degree of dogma and assumed that by now it would all have been straightened out. It hasn't.

Due to the inertia of the Copenhagen Interpretation of quantum mechanics, seeking the foundations of physics was not deemed a "respectable" pursuit until after 1970 which roughly coincided with the first issue of a journal called "Foundations of Physics" [Found]. This was a new opening for publications with a floodgate of stored-up criticisms, speculations, reconsiderations, inquiries, alternative views and philosophies and was a liberation to me—the arena I most cared about.

I've taken more than a hundred classes, but most of my "real" learning has been on my own outside of classes. I loved teaching courses but would prefer never to take another class under someone else's direction. I want my own intuition to guide me towards what is really fundamental. I am "launched," so I can do it myself. Since 1960, I was a regular patron of our university math-physics library and spent a lot of time and effort searching for basic answers in journals and textbooks. Now I can ask Google ten questions a day and get to see physics articles daily on the web (such as ArXiv.org, Physics World, Science News and physics.aps.org). Promising web articles get Xeroxed daily for study, and key information is then documented for possible use in my own essays. There is a tedious process of <u>gleaning</u> for little bits of leading information – hoping that someone somewhere someday will make a statement that "spills the beans" and divulges his own precious perspective.

Two such examples that inspired me were:

Claim: *For experiments on electrons and photons, "*<u>*There are no particles*</u> *…there are only field quanta--excitations in spatially extended continuous fields." [Art Hobson]*

An elementary “particle” is something of a deduction after interaction, collapse or measurement. Perhaps the term “quanta” should be kept distinct from composite-particles like the proton or a “buckyball” C_{60}. But, the “matter-wave” ψ is still common to all. At least we are no longer talking much about an electron “particle” traveling simultaneously through two slits. Perhaps it is more like a soliton wave guided by a ψ-wave.

Another inspiration is the quote:

"But our present QM formalism is not purely epistemological; it is a peculiar mixture describing in part realities of Nature, in part incomplete human information about Nature --- all scrambled up by Heisenberg and Bohr into an omelet that nobody has seen how to unscramble." [E. T. Jaynes]

So, rather than the perpetual “frog-mouse-battle” between either “our knowledge” versus “existing reality,” we may instead have “both/and” with a mixture that may vary from case to case. For example, after being free from the nucleus, an alpha emission decaying from U-238 may really be an α-ray trajectory. An observer assembles an ensemble of α-detection data and deduces, “Oh, its wave function clearly has spherical symmetry and must be like $\varphi = Ae^{ik\cdot r}/r$.” A traditionalist might say, “it is a spherical wave until there is a first ionization maybe in a cloud chamber, and that constitutes a ‘measurement’ that collapses its wave-function into a ray.” But is that really true? Another possibility is that it “really” is a particle-ray but only after it knows about its future detection (quantum information “Back from the Future”).

Learning physics and seeking its foundations is an intellectual pursuit. That means constantly reading, constantly learning, trying to be efficient, working hard, having deep curiosity, valuing truth, being suspicious of dogma and authority, appreciating ideas, generally avoiding the small talk and shallow socializing of standard life, enjoying being alone, and sacrificing for the future. A goal is to form huge dovetailed mental structures of ideas and mathematics ideally based on some small number of postulates.

Further Thoughts on Topics in Modern Physics:

“Why the Quantum?” is an apparently permanent mystery in quantum mechanics. de Broglie’s rules of 1924 are $E = \hbar\omega$ {also called the Planck-Einstein relation} and $p = h/\lambda = \hbar k$ for massive particles as well as photons. I addressed the meaning of this in my final essay, “An electron is waves of what?” de Broglie began with the concept that all microscopic mass/energy vibrates: $\omega = E/\hbar = \gamma mc^2/\hbar$ and that $p=h/\lambda$ is merely a Lorentz transformation of that vibrating mass (essentially a relativistic “clock de-synchronization” viewed by an observer in relative motion). Schrödinger kept the $p=h/\lambda$ rule but only used kinetic energy for $E+V = \hbar\omega$ while ignoring the rest mass. Psi, ψ(x,t), is a ψ-wave of an information-bearing ψ-field: analog energy is deduced as the density of waves in time, and momentum is the density of waves in space. ψ itself is not a wave of any kind of energy, it is more like a “quantum energy amplitude” with a blurry interpretation due to the concepts of “collapse” and the Born Rule. The ψ-wave only communicates “mechanical” properties of E, p and angular momentum; and its wave-like “amplitude” can interfere with itself.

What is fascinating is that the equation $p = mv = h/\lambda$ applies to almost “everything:” elementary fermion particle motion (with mass from the Higgs field), protons (which are mainly confined energy, $m_p = E/c^2$), phonons (mechanical vibrations), massless photons (electromagnetism), and very large moving molecules with extremely tiny wavelengths, $\lambda = h/Mv$. de Broglie waves are completely generic and not made out of anything classical that we know. We have to accept that these are fundamental rules imposed and maintained by Nature, "Whatever it is that forms energy, it is constrained and governed by an 'energy supervisor' that controls the packaging of quanta, the shipping of quanta, and enforcement of the conservation of energy." Is the explanation of this rule discoverable or hidden inside the "Core?"

Fields are the most important things in quantum field theory (QFT). Our simplest view is that they are like a 3D "mattress of springs" filling spacetime. Traveling waves are sequences of compression and expansions of the little springs. Particles and energy levels are quantum excitations of its normal modes. There is an occasional belief that an electron wave being a disturbance of a special electron quantum field might imply that an electron quanta could be reconstructed from a de Broglie wave as part of a collapse at detection. But, there is the illuminating experimental fact that a massive spread-out molecule composed of thousands of atoms also has a simple matter wave, and that encourages belief in particles remaining and traveling as particles subject to guiding waves—a partial return to de Broglie-Bohm views. If that is true for big particles, then for consistency, why not also for the electron? For oscillating elementary fermion particles, we might view the particles as confined (soliton-like) energy waves producing extended information or guiding waves beyond the particle –in part a return to de Broglie's original but incompletely developed "double solution" model. A local particle "u-field" and broader ψ-field both vibrate at the same frequency. Is it possible that these waves are related to the phenomenon of jitter-motion "Zitterbewegung?" – it does pertain to fermions, massive and massless scalar bosons, and Proca type spin-1 bosons too – but its frequency is double that of de Broglie (it contains terms with the factor e^{-2iHt}).

The increasing number of different proposed interpretations of QM has expanded the set of possible assumptions that might go into a new interpretation. Photons are their own antiparticles and might be able to go backwards in time as well as forwards. The possibility of Cramer's vaguely defined "pseudo-time" makes an emitter-absorber transaction an evolving "process" rather than a single world line. Aharonov's two state solution has a final state advancing backwards in time. Feynman also allowed electrons to also go backwards in time as contributions to his path integral. And we already mentioned the de Broglie "double solution," having both a real-physical u-wave localized close to a "particle" along with a ψ-function perhaps representing our knowledge of probable outcomes over an ensemble of events.

Quantum Mechanics:

It is often difficult for layman and non-specialists to access and grasp knowledge of important fundamentals of a given subject. Texts don't like to admit that there are important concepts that are still unclear or unknown – authors don't like to state their ignorance. The chemical covalent bond for example is a quantum effect. Plugging and grinding a detailed Schrödinger equation for some appropriate "Hamiltonian" can approximate the right numerical answers. But the basic acting principles for intuitive understandings are not obvious and are rarely offered in chemistry books or even in physical chemistry books.

What could be more important than knowing what holds atoms together in molecules? All chemistry books show bold "Lewis dots" {for example H **:** H } of electrons lying between and being shared by atoms – a pretty poor and unrealistic picture begging for more clarity that isn't offered. An intuitive component of binding is the "enhancement of effective electron charge density between two nuclei." Orbital overlaps get squared by the Born rule thereby promoting quantum amplitudes to an increased electron density greater than that of mere linear superpositions. Many chemistry texts don't even mention the Born rule! – but they do picture "charge densities" around atoms. And, there is an "obvious" but also rarely mentioned contribution: electrons in molecules share several nuclei and hence have a bigger play-room in which to move. That broadens their de Broglie wavelengths and lowers their effective kinetic energy. These are big players, but there are others even more difficult to reveal. There may be some mention of a need to add a little "ionic character," include a fraction of 2s and 2p orbitals, and make the radial size of the 1S-orbitals variable and constricted (more localized). Interpretation is tricky in chemistry.

Chemistry texts don't mention these principles because quantum interpretation is unclear even in basic chemistry. I was once a "boy chemist" but got turned-off by high-school memorizations without understandings.

My biggest personal hunch is that the "psi" in quantum mechanics represents the "square root of reality" [in ways that my FQXi essays have discussed]. This description and slogan is one of my favorite analogies but doesn't appear to be favored or even mentioned in any reference. As discussed before, Dirac's biggest discovery was that Dirac fermions are effectively governed by "the square root of the relativistic Klein-Gordon equation" {and that <u>is</u> mentioned in references}. In general, the square root concept forces us to dig deeper into "Clifford Algebras" which generalize hypercomplex numbers and beyond. Physical reality is somehow isomorphic to the actions of these algebras. That means that there is a real mapping between the math and the physics that the math represents. For me, the beginning of this realization was Cramer's transactional interpretation of QM involving "retrocausality" which also seems to be required for understanding entanglement and may be also the only sensible explanation for the Born rule.

As examples of "square-rooting" *(or conjugate-star-rooting)* a simple popular *{but perhaps over-simplified}* expression for the "wave function of a photon" is the Riemann-Silberstein form $\psi^*\psi \propto (\epsilon_o/2)(E^2+c^2B^2)$ energy $\rightarrow \psi \sim (\epsilon_o/2)^{1/2}(E \pm icB)$ resulting in complex numbers. Unlike the scalar field de Broglie waves, E and B are vectors that can have polarizations.

But, as mentioned before, a most important example is the **<u>Dirac</u> equation,** $(i\,\partial\!\!\!/ - m)\psi = 0$, (short for $i\hbar\,\gamma^\mu\,\partial_\mu\psi - mc\psi = 0$) which is the "square root" of the "Klein-Gordon" equation, $[\partial^2/\partial(ct)^2 - \nabla^2 + (mc/\hbar)^2]\,\psi = 0$. {This may also be considered as taking the "square root" of the d'Alembertian operator, $\Box$ {the term "square root" meant something like deducing possible ψ's when given $\psi^*\psi = |\psi|^2 \in$ Reals }.{Feynman "slash notation" $\partial\!\!\!/$ has a slash through the partial derivative standing for $\gamma^\mu\partial_\mu = \partial\!\!\!/$ {slash partial }, called the <u>"Dirac Operator,"</u> and requiring the use of hypercomplex gamma matrices where $\gamma^\mu\gamma^\mu = \pm I_4$}, and $\Box = \partial\!\!\!/\,\partial\!\!\!/$. Or, symbolically, the Dirac Equation = $(KG)^{*/2} = (i\,\partial\!\!\!/ - m)\psi = 0$ (where "natural units" are $c = \hbar = 1$). }

The orthodox meaning of ψ in Copenhagen quantum mechanics is "waves of probability amplitude;" an interpretation that is familiar but also numbingly opaque. The name is certainly appropriate "for all practical purposes" *[Bell's acronym is "FAPP"].* But that need not be the whole story. Matter waves could be "quantum-real" prior to measurement but eventually couple to a separate last stage action "Principle of random selection of absorber" or "stochastic choice" of ψ*ψ "intensity presence" -- a two-step process. A familiar analogy is physical electric waves with amplitude E(x,t) finally materializing as detected photons based on their space-time energy density $\propto E^2$.

ENTANGLEMENT

In apparent violation of causality and relativity, two particles are entangled when the quantum state of each particle cannot be described independently of the state of the others.

An example is a wavefunction for particles a and b described in configuration space by:

$$\Psi(r_1, r_2, t) = \psi_a(r_1,t)\ \psi_b(r_2,t) \pm \psi_a(r_2,t)\ \psi_b(r_1,t)$$

Where + is used for symmetrical wavefunctions and "–" for anti-symmetric ones {meaning 180^o out of phase}. If particle "a" is found at location r_2, then b has to be at r_1. Or, for spins, if a is measured to have one spin, s_1, then immediately the spin on the other particle takes on its required spin, s_2 {and similarly for light polarizations}.

Counter to human intuition, experiments with jointly entangled particles show that these correlated outcomes have no dependence on event times, spatial separation or order of events as if some sort of long range back-and-forth-in-time "sub-quantum communication" occurs between entangled particles. This "spooky action at a distance" is like a magic act that is continually performed right in front of us – and we just can't see the "trick."

John Bell showed that the entanglement correlations revealed by experiment are not pre-determined by some previous contact — they occur spontaneously and non-locally {or much faster than the speed of light}. He devised some test inequalities that have been experimentally verified over and over again. One of my papers is about what correlations could result from pre-determined Concrete Hidden Variables for the purpose of comparison against actual Bell-test experimental results and then dismissing them in favor of standard quantum mechanics. We can elaborate on these statements as follows:

Entangled particle "Bell" test results are usually given just as inequalities and usually for photon experiments with polarizers. This should be compared to what would happen for the more classically intuitive but incorrect case of particle behavior being predetermined from some common initial source. This is what I call the "concrete" case, but few students are familiar with how hidden variables would work for these cases. These concrete examples are generally not derived or included in Bell test papers.

In contrast, an actual quantum mechanical test of one photon of a pair immediately "snaps" the other photon into alignment. To appreciate this, one must be familiar with some scenarios incorporating predetermination, and that is the purpose of this paper. Hidden variables have contrived or classical mechanisms that might approximate a similar output , and its calculations often apply "convolution integrals." Actual test statistics are then compared to these concrete thought tests and found to be in disagreement as Bell claimed.

It has occasionally been suggested that the quantum entanglement correlations may be due to back-and-forth communications in time that emulate instantaneous communication between measuring systems. The "no-signaling" theorem of quantum mechanics firmly says that no bits of classical-information can be transmitted back-and-forth in time. But, a possible solution is to suggest that *"quantum information"* operates in a different sub-quantum-realm, and transmission can be bi-directional in time. This has been called "retro-causal," but the term causality refers to cause-and-effect relations in the classical world.

There are now a vast number of entanglement experiments that disturb traditional thinking and give the "appearance" of some sort of retrocausal communication. This is an especially tempting solution for the myriad "delayed choice quantum eraser" experiments and "entanglement swapping" lab tests. The most interesting of these is "entanglement between photons that have never coexisted !" [Meg]. The earliest suggestion of a back-in-time explanation for entanglement correlations seems to have been by Costa de Beauregard in 1953 and has been referred to as the "Parisian Zig-Zag."

{My paper, "Appearances of Retrocausality" with "Entanglement Figures"}

WORDS and The Possibilist Reality:

"The limits of my language are the limits of my world" [Ludwig Wittgenstein].

In the social, economic, legal, military, and political worlds, unbiased "truths" may not be highly valued, and doublespeak is frequent. Examples are: "enhanced interrogation" {torture}, department of defense {offence}, "states rights" {race control}, neutralize {kill}, intelligent design {god did it}, pacification {bombing}, ethnic cleansing {genocide}, corrections {prisons},…). History also tends to be biased and even mythological because it is justified and written by the "winners." And there are countless examples of outright lies such as the year-after-year claims over nearly two decades that "We are winning the war in Vietnam," "We are winning the war in Afghanistan."

In contrast, physics and science in general cares deeply about truth and dislikes obfuscation. But there does exist some degree of "fashion" or paradigm in physics. As fashion changes, older concepts and words have an inertia that may not dovetail with newer needs; and their continued use can lead to confusions. With poorly defined or obsolete terms, scientists may talk with cross-purposes without understanding each other.

Neils Bohr and later-on Werner Heisenberg set the fashion for the first half century of the quantum mechanics from 1926. Heisenberg's positivistic paradigm, which in 1955 he labeled as the "Copenhagen Interpretation," said that we were not supposed to talk about a quantum subworld but only about its pragmatic results in the classically measured world. Few believed that the wave function represented something "real." In 1926, Erwin Schrödinger was dismayed to realize that his N-particle wavefunctions propagated in an abstract fictitious "configuration space" of dimension 3^N. This concept was quickly accepted but made it difficult to believe in the reality of such waves unless the word "wave" was strongly broadened beyond our usual understanding [Afriat]. Unlike the classical case, an N-particle quantum wavefunction cannot be described in our ordinary space. Having a ψ that was complex and often lacking in any absolute phase also seemed unreal.

We all know that the quantum world is not classical and should require its own new language to describe the processings of the wave function as a different "reality." Discussions often used old terms, but in parentheses to mean something perhaps different but related (like "reality" or "wave" or "particle"). If we wish to discuss the world explored by ψ prior to its "collapse," there presently are no clear and accepted words. The processing of "possibilities' may be best, and I've previously used "QuReal" or "sub-classical," or psi-real in "psiland." $\psi(x,t)$ is governed by differential equations such as Schrödinger's, but ψ won't do anything until it explores its boundary conditions (BC).

<u>The world of alternative possibilities as an example of a sub-quantum reality.</u>

In particle physics, there are often a number of different interactions or decay choices that could occur; and the different possibilities can interfere with each other.

For example, a charged pion can decay in three possible ways of which only one is finally chosen: $\pi^+ \rightarrow \mu^+ + \nu_\mu$, or $\pi^+ \rightarrow e^+ + \nu_e$, or even $\pi^+ \rightarrow \pi^0 + e^+ + \nu_e$.

In ordinary quantum mechanics, a photon that is produced by a quantum transition might travel in many possible directions of which one is selected and detected according to the probability density $\psi^*\psi$. The transactional interpretation of quantum mechanics ("TI") considers the wavefunction $\psi(x,t)$ to be an "offer wave" ("OW") from an emitter to a number of possible absorbers which in turn produce "confirmation waves" ("CW") <u>backwards-in-time</u> to the emitter. There is a transaction or "hand-shaking-agreement" between an emitter and one of the absorbers enabling the physical transfer of the photon. Quantum physics has always considered an actual event selection to be purely random – but perhaps it is emergent out of complexity and so is only apparently indeterministic.

I've liked this interpretation ever since John Cramer proposed it in 1986 [Cramer]. It is the only interpretation that intuitively gives the Born Rule and in my mind also explains entanglements. Its main advocate now is the physicist/philosopher Ruth Kastner who calls it "Possibilist TI." She "argues that OWs and CWs are possibilities that are real" -- "less real than actual empirically measurable events, but more real than an idea or concept in a person's mind" and suggests the alternate term "potentia" [Kastner]. Lee Smolin says that the "world of the possible has to be included as part of reality—because in quantum physics the possible influences the future of the actual."

<u>PSI EXPLORES THE REALM OF POSSIBILITIES FOR EACH FINAL EVENT.</u>

I believe that there is a "reality" below classical reality – but it requires a broadening of the word "reality." This is a world we so far cannot directly explore, but we can deduce that the quantum psi engages in <u>active exploration</u> prior to "collapse" (the Purcell effect is an example). It behaves almost as if it had some intelligence behind every event. As an analogy, there is a "process" in which the wave-function first "cases out the joint" before robbing the place. Every real quantum event has sampled its own world of possibilities prior to the action of a "principle of random absorber selection. " The possible future aids in actualizing the past. The process, whatever it turns out to be, is the sub-reality.

Edward Purcell noted that an excited atom placed in a small reflective cavity can have a very different decay lifetime that can be strongly enhanced or even totally suppressed from that in "free space." Since 1946, this deviation from a free non–interacting half-life we now call the "Purcell effect," and it was finally demonstrated at optical frequencies in 1987. It is very interesting that this effect can also occur over quite long distances [Herzog {-- this amazing –but overly concise—article from 1994 is discussed in my paper on "Appearances of Retro-Causality in Entanglement Experiments"}]. An atom doesn't emit until it knows its environment.

The interior of a highly reflecting cavity only allows wavelengths that fit ($\lambda/2$, 1λ, $3\lambda/2$, ...). If an atom is placed in a small cavity having a size smaller than the transition $\lambda/2$, no photons can propagate and the atom is unable to decay at all. *In a cavity at resonance, the density of final states is enhanced.* This is a result from cavity quantum electrodynamics (CQED) based on Fermi's golden rule which dictates that the transition rate for the atom-vacuum (or atom-cavity) system is proportional to the density of final possible states.

In symbols this is: $\Gamma_{i\to f} = (2\pi/\hbar)|\langle f|H'|i\rangle|^2 \rho(E_f)$ where Γ is the transition or decay probability for an initial state $|i\rangle$ and final state $|f\rangle$ and ρ is the density of final states near E_f. Then the probability of $|f\rangle \propto e^{-\Gamma t}$. Quantum optics uses a density ρ of photon states near $\hbar\omega$.

RELATIVITY

The invariance of the speed of light, c, is essentially an "axiom" of special relativity; and consequently all physical theories are supposed to be Lorentz invariant. But, Why? The speed limit is perhaps a clue about how fast any kind of information can travel and must have some origin in the fundamentals of the Vacuum. It tells us something basic about space-time that we have not yet fully grasped.

The metric of relativity goes beyond familiar 3D Euclidean distance to the "proper time" between events (4D space-time pairs of points). If two events are connected by the speed of light $\Delta x/\Delta t = c$, then proper-time or distance is zero!! I take this strange concept to be fundamental, and it should be more fully exploited. Also note that entanglements seem to say that "distance doesn't matter" and they entanglement tests most often use light photons at speed c.

Restating this, relativity uses an initially strange "pseudo-Riemannian" metric. What should be initially presented is that distance in 4-dimensional relativity is changed to $ds^2=c^2d\tau^2$ where the important "tau" is now the proper time. Events connected by light have no metric separation so that $c^2d\tau^2 = c^2dt^2-dx^2$ gives zero change in proper time. Dividing this metric by c^2 then gives $d\tau = dt\sqrt{(1-v^2/c^2)} = dt/\gamma$, or $dt=\gamma d\tau$ (the usual time dilation formula—quick and easy). If v=c, then $d\tau = 0$ (zero 4-distance).

Why is the metric of special relativity so concerned with measuring "deviations from the speed of light" as the "4-distance" measure called "proper time?" And why do entanglements not care about distance? There was an instant in the early universe when there were no deviations, no mass, no non-relativistic quantum mechanics and no photons. We owe our physics to the transition called "electro-weak-symmetry-breaking" [EWSB]. Is the next paragraph at all relevant?

Quantum Mechanics emerged in an early period of the Universe when everything was up close and personal (essentially no "distance" separations). Everything was at the speed of light—all connections were light-like – but that doesn't mean light as we think we know it. The photon epoch began at 10 seconds after antimatter annihilation. But the first emergence of photons was at EWSB at 10^{-12} seconds (about a pico-second). Prior to that time, there were no photons nor any massive particles. There were only massless fields. So, did quantum mechanics exist then? Is what we call mass the same as frequency of zig-zags in the Higgs field? Unfortunately, EWSB still lies "beyond the standard model," so it is not well understood.

In my paper on "Learning quantum mechanics and relativity," I note that the single particle Lagrangians for both classical physics ($L = T - V$) and special relativity ($L = -m_o c^2/\gamma - V$) might be better presented as being proportional to "wave counters" – counting waves or phase along a path for a "path integral." That interpretation gives the abstract "Lagrangian" more tangibility. It is interesting that it carries over to classical mechanics without that apparent need.

In general relativity, I took another look at how black holes are formed according to the historical 1939 "Oppenheimer-Snyder" or "frozen star" formulation {the essay here is "Brief Summary of Collapse to a Schwarzschild Black Hole"}. Coordinate time is really frozen near the horizon ($g_{oo} \rightarrow 0$ and $g_{rr} \rightarrow \infty$) -- so we will never see a particle actually penetrate the horizon. Sparse falling matter or inwards-directed light will be seen to accumulate just outside the evolved Event Horizon. The cycloid equations of a dynamic closed cosmology are also useful in describing this collapse.

Unfortunately, all black holes rotate; and going deeper into the math-physics of these black holes is very difficult. It took almost fifty years of hard thinking in differential geometry to finally progress from the static Schwarzschild solution of Einstein's 1915 equations of general relativity to the rotating Kerr metric of 1963.

On the cosmology front Recent local Hubble estimates indicate that the universe is growing ~10% faster than indicated from analysis of the cosmic microwave background radiation (CMB) [Riess]. A few astrophysicists are referring to the disparity of universe expansion rates as a "crisis." Some adjustments might be needed in the standard six-parameter ΛCDM model (dark-energy with cold-dark-matter model).

PARTICLE PHYSICS

The modern physical Vacuum is "not-nothing" but rather the seat of fundamental physics (physical laws, constants, physical objects such as "elementary particles" as unique and omnipresent "templates" in the Vacuum). The Vacuum contains fields for each elementary particle.

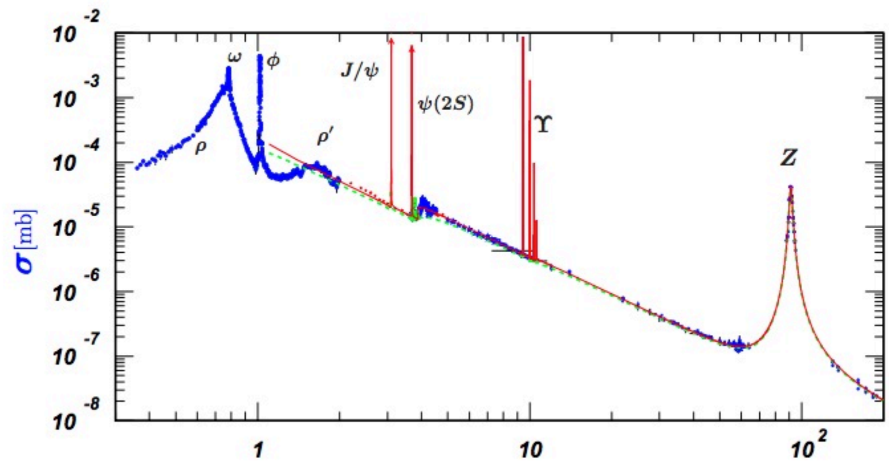


Figure 1. The cross section, σ, for hadron production from electron positron annihilations versus √s energy in GeV {total energy in the center of mass frame, "s" is "space-channel"}. [Particle Data Group, http://pdg.lbl.gov/2007/reviews/hadronicrpp.pdf].

With a supply of local energy, elementary particle pairs {quark-anti-quark mesons} can just **"pop out of the Vacuum"** of space-time [Figure 1]. Lepton colliders for e^-e^+ or $\mu^-\mu^+$ annihilate into photons. As their energy is increased in giga-electron-volts, we first emerge quark-meson combinations of $u\,\bar{u}$ & $d\,\bar{d}$ quarks called ρ or ω mesons. This is followed by $s\,\bar{s}$ or ϕ mesons, then charmonium $c\,\bar{c}$ or ψ mesons and $b\,\bar{b}$ = ϒ (upsilon) mesons {t, or top-quarks, don't live long enough to actualize}. Finally we see Z^o particles (the neutral weak boson or "heavy photon"). The thin vertical spikes show that the incoming energy has to be precisely on target to stimulate the heavy meson resonances of the various quantum fields.
{The ρ' is an excited state of ρ, and the ψ(2S) is an excited state of the J/ψ (1S) }.

All of the particular fermions (quarks and leptons) and bosons are "identical particles" **as if** their unique templates pre-exist in the Vacuum. Identicalness doesn't exist in classical physics but is rather a feature of quantum field theory, QFT.

Since pumping energy into a point in the Vacuum produces particle pairs such as quarks plus anti-quarks or muons plus anti-muons, it would appear that these elementary particles are actuated resonances of the underlying structure of the Vacuum.

Why did I call the Vacuum "Plato's Form Heaven?" Because the essence of his Forms or "Ideas" is invariance, particular things or qualities taken from universal templates (circles, spheres, trees, color blue, gold, beauty). If Plato were alive today and knew that identical electrons were duplicated 10^{80} times throughout the universe as quanta from the electron field – he certainly would call electron-ness a perfect invariant Form from immanent vacuum fields. All elementary particles are Forms as also is most of the constructs of mathematics.

The mechanisms that function at our familiar sizes depend on reductionist mechanisms at very tiny scales – how deep down have we seen Nature go? A related question is how much energy a particle can have. In the 1960's, we were discussing energies in MeV and GeV but not in TeV nor PeV nor EeV's. But now we are measuring gamma ray and neutrino and cosmic ray energies in these upper realms and beyond [e.g., 10^{20} eV!, Halzen]. Is there a limit?

In the realm of "beyond the standard model" (BSM), "string theory" (a misuse of the word "theory") is an interesting but also dangerous arena of advanced physics that might be forever beyond testability. If it can't be tested, is it still science? Should the public respect claims without proof? Supersymmetric string theory has long predicted that there would be testable consequences for experiments at the Large Hadron Collider (LHC) at CERN. But, the declared final result of high-statistics 13-TeV high energy proton testing there was that there was "no evidence at all" for super-particles. Nevertheless, string theory and the "multiverse" and holography are still often thrust upon the general public as if they were all real physics – and a recent 3 million dollar prize was just awarded for supergravity (the Milnor Breakthrough prize). I would agree with Woit, Hossenfelder, and Smollin that this should presently be more in the realm of some rational religion rather than science.

Nevertheless, I very much like the idea of higher dimensions as providing a needed substrate justified by the great complexity of known field theories (e.g., like the Kaluza-Klein paper in Book 1 providing an extra dimension to express electro-magnetism). And I would like to think that we might someday agree on an interpretation just based on agreed most probable deductions from known behavior and carry that over to quantum mechanics as well. It may be true that "beneath quantum mechanics" is also untestable. But I believe that some such reality exists nonetheless and will eventually be supported (in some way).

A paper "Lie Group Representations" outlines Gel-Mann's 1961 hierarchies of the baryons and mesons. The important case of the spin 3/2's "decuplet" **10** can be constructed simply and intuitively since all quark spins are aligned and the states are completely favor symmetric. These hadrons include the Δ's, Σ's, Ξ's, and the famous omega-minus strange particle Ω^- (sss) whose confirmation first made the quark idea respectable in 1964. The Δ^+ and Δ^o are like excited states of the proton and neutron.

VIEW ON MATH IN PHYSICS

Mathematics is the language of physics, and It is interesting to see just how far mathematical thought has gone in describing and evolving each field of physics. As Feynman said, math is not just another language, it is a language with reasoning built into it. Eugene Wigner once wrote a popular paper called "The unreasonable effectiveness of mathematics in the natural sciences" [Wigner_1960] that said:

"The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve." And, "the mathematical structure of a physical theory often points the way to further advances in that theory and even to empirical predictions."

I based my FQXi essay, "Physics lives in Form Heaven" on this statement.

There are special examples of how learning a new field of mathematics grants a student a new level of power, understanding, and perspective not previously available. The most obvious example, of course, is the great power of calculus towards enabling an understanding of classical physics.

A lesser example is realizing that the huge ranges of size and space can only be conceived by viewing the world logarithmically – the log of durations of time versus log of sizes of space that is also called a "powers of ten" view as expressed in our "scientific notation."

"Connections" are basic in general relativity (the Γ symbols) and in general in differential geometry. Einstein would not have completed his goal without the use of tensor calculus. The concept of connections as consequences of transporting data along a curve can be made intuitively clear by first considering simple examples such as moving along a 40^o latitude on a sphere (place a dunces-hat cone on that latitude).

And consider the mathematics of coordinate invariant <u>"differential forms"</u> such as "1-forms" α or 2-forms ω. In a loose sense, forms are just things that can go under an integral sign – like say $\int 2x^2y\,dxdy = \int \omega$. We can treat the dx and dy as base vectors in ω and add a seemingly trivial "antisymmetry rule wedge product" that just says that order counts: $dxdy = dx\wedge dy = -\,dy\wedge dx = -dydx$, and $dx\wedge dx = 0$. This little rule has remarkable consequences.

Forms also incorporate an "exterior" differential operator **d** with some of its own remarkable powers (similar to $d\mathbf{r}\cdot\nabla\wedge$ -- as an operator) . An example is electromagnetism from the scalar potential ϕ and vector potential A. Certain derivatives of ϕ and A yield the fields E and B, and the real world of energy acknowledges the physical existence of energy densities E^2 and B^2 . One might conclude that the space-time Vacuum must actively process potentials by differentiation into EM fields. The differential operator **d** of a 3-space "vector" A-field yields the magnetic field pseudo-vector "B" = **d**A.

It is amazing that in Minkowski 4-space we can simply write **F** = **d**A where A is now a "4-potential" (ϕ, A) but written as $A = A^o\,dt + A^1dx^1 + A^2dx^2 + A^3dx^3$ {see paper on "Geometry in Modern Physics"}. **F** is the anti-symmetric **Faraday** electromagnetic tensor of special relativity ($\sim F_{\mu\nu}$). That is, "**d**" can make derivatives and curls and also the generalized 4-curls that produce **F**. Of all mathematical languages, the concise power of this derivation leads one to wonder if this economical language of forms is <u>optimally isomorphic</u> to the processings of Nature's Vacuum. There are mappings between mathematical physics and physical reality, and some mappings may be more "real" than others.

Furthermore, **d**F = **dd**A = 0, which happens to yield "Faraday's Law" ($\nabla\times E = -\partial B/\partial t$) along with Maxwell's "no poles" ($\nabla\cdot B = 0$). These observations apply to the EM gauge group U(1). If we consider the higher gauge group SU(2) for say Yang-Mills theory in QFT, then there are analogs to the familiar A and F that have a more general curvature 2-form: $F = dA + A\wedge A$. { Since U(1) is Abelian, it's $A\wedge A = 0$ – -but not for SU(2) }.

Spinors:

Spinors are objects that somewhat resemble taking the square root of vectors and are usually pictured as short column vectors. See paper on "Spinors."

Roger Penrose intuitively defined a spinor as an object which turns into its negative after a complete $2\pi = 360^o$ rotation. An example could be the Mobius band where <u>two full rotations</u> are needed to get back to the original orientation. He adds that the action of rotation on a spinor is always double valued, 2:1. Beginning quantum mechanics shows Pauli matrices operating on 2-spinors as complex fractions of up and down electron spins. Broader examples include matrices representing continuous ("Lie") groups to transform spinor column vectors. Group names use special symbols like U for "unitary" {where $(U^T)^* = U^{-1}$ }, "S" for "special" {meaning having determinant = +1}, O for orthogonal where $A^T = A^{-1}$, L for linear invertable, and C means complex.

Consider the Lie group 2:1 homeomorphism for the 2×2 complex matrices of SU(2)→ SO(3) {the group of rotations in 3-space} and also the case of the special linear group SL(2,C)→ SO(1,3) for 4×4 Lorentz transformations {the (1,3) refers to 1 time dimension and 3 space dimensions of a Minkowski metric}. Two elements of SU(2) map to the same rotation element of SO(3), and this is called "double covering." Two-component spinors transform via multiplication by elements u∈ SU(2), and elements of u contain "half-angles" $e^{-i\theta/2}$ which get imparted to the spinors (to give the needed "twice arounds"). The elements of the Lie algebra su(2) {the "tangent" space of SU(2)} are intended to represent hypercomplex quaternion vectors. Dirac's standard 4-spinors are mixtures of "Weyl chiral" 2-spinors}.

Electron spinors are can be viewed as hypercomplex quaternions {H for Hamilton}, Pauli matrices are ∈ C × H, and Dirac matrices are hypercomplex "Clifford" entities.

Solid State/ Condensed Matter Physics:

In 1982, David Thouless published a paper called "Quantized Hall conductance in a two-dimensional periodic potential" with research fellows Kohmoto, Nightingale and den Nijs (labeled as "TKN2") and was later cited for the Nobel Prize in Physics. The word topology is not mentioned in the title of the 1982 paper and does not appear in his titles until 1985—but it is there implicitly.

The introduction of topology into the behavior of exotic solid state materials is now a major field of condensed matter physics. The first and best-known example is the "Integer Quantum Hall Effect" ("IQHE" or just "QHE") for a two-dimensional electron gas surface. A 3d bulk insulator can have a 2d conducting surface with 1d edge currents. The role of topology in materials often enters through quasi-momentum on the "Brillouin torus" for crystal lattices with geometry that repeats from atom to atom. That is, a periodic crystal surface (say cubic) may have matching periodic wavefunctions throughout. Identifying opposite sides of a rectangular cell effectively results in a torus topology -- the Brillouin torus.

Beyond the IQHE phenomenon is the "Fractional Quantum Hall Effect" (FQH). Both of these topics have led to Nobel Prizes in physics. All of this is discussed in my paper on the "Nobel Prize for Topology in Exotic Materials."

IQHE is now as significant as superconductivity. In strong magnetic fields (e.g., 10 Teslas), the Fermi levels on the 2D surface have integral numbers of wavelengths. So, Bohr orbitals didn't work in 3D, but the idea does apply in 2D.

To the general public, topology describe holes in a manifold {a donut is like a coffee cup in that each has one hole}. But having a 2D semi-conductive surface on a 3D insulator and 1D edge currents is also topological.

MY FAVORITE HIGH-TECH STUDY PROJECT:

From 1965 to 1972, I started three different PhD graduate programs but never completed a thesis. One project of 1967 was to show that the Carbon-12 nucleus contained three α particles by processing data supplied from a "pions on propane" experiment. Now, after 50 years, the model that "within a nucleus there are substructures of alpha particles… has not been accepted despite the considerable evidence for its validity" [Watkins]. {Implying that it wouldn't have been a very productive project}.

In my later high-tech career, one of my projects, a theoretical versus experimental study of data disk life, could have served as an acceptable thesis.

Since about 1980, magnetic recording hard-disk data drives have required thin surface lubrication as a protection against high RPM "head crashes" which would destroy the magnetic surface. My study of long term flow of these thin lubricants on the surfaces of spinning disks involved a decade of mathematical modeling, measurement and modifications of disk surface porosity to finally ensure a long product lifetime.

It had been a common belief that radial lubricant migration was caused by centrifugal force from rapidly spinning disks. In a variety of approaches, I showed that it was instead mainly due to the effects of <u>air-flow wind shear stress</u> on the very thin surface layers of lubricant; and for the thin viscous films on particulate disks, this effect was much stronger than inertial forces. Depleted lubricant can be partially replenished from lubricant stored in the porosity of the disk binder. Boundary layer fluid theory was applied in my modeling.

I gathered three sets of long term measurements of lubricant thickness profiles over time: ESCA for lube on the surface, FTIR for total thickness lubricant, and chemical Freon "strip-and-weighs" by annular sections across the disks. All three methods dovetailed and agreed with the final models. A newly encountered phenomenon was complete depletion on inner diameters due to "polymer slip" from the long serpentine poly-fluoro-ether molecules, and this was included in the model.

Unfortunately, what is current in high-technology has a very short half-life. Next generations make older generations irrelevant. <u>One of the best things about basic science is that it lasts forever.</u>

Physics largely involves problem solving and mathematical modeling. And, regardless of title, that is essentially what I did over my whole career. Out of a thousand cases, I include an elementary model here for the writing of alternating transitions on a moving magnetic medium: "MagWave." I liked it, it was pretty. One interesting aspect of High-Tech problems is that they are unique, generally never solved before. My approach was to struggle with a problem as a way to "load it into my RAM" brain memory and then sleep on it. At night, my little brain micro-processors would work on the problem and partly solve it. Then in the morning, they flash their solution on my internal conscious vision screen. I don't know if that is unusual or not—but it worked well for thirty years.

Every month, I also put out a list of about ten significant discoveries taken from readings of major literature sources (not all science, physics and math). I include a set of these at the end. For example, a recent summary note was:

∘ *There are now "an astonishingly high number of black holes of all types in the contemporary and early z~10 universe. Practically all black holes in the universe are primordial PBHs (this defies most accepted models). There are also new quasars with z> 6. Perhaps BH's came first and galaxies followed [from ArXiv.org 1911.023382].*

References:

[Wig_1960] Wigner, E. P. (1960). "The unreasonable effectiveness of mathematics in the natural sciences." Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959". Communications on Pure and Applied Mathematics. 13: 1–14. And also https://en.wikipedia.org/wiki/The_Unreasonable_Effectiveness_of_Mathematics_in_the_Natural_Sciences

[Halzen] Francis Halzen, The Highest Energy Neutrinos, https://iopscience.iop.org/article/10.1088/1742-6596/120/6/062004/pdf

[Thouless] David J. **Thouless** ,"Quantized Hall conductance as a topological invariant. {David passed away on 6 April 2019}. https://www.nobelprize.org/prizes/physics/2016/thouless/biographical/

[Smolin] Lee Smolin, Einstein's Unfinished Revolution. The Search for What Lies Beyond the Quantum, 2019.

[Cramer] John Cramer, "The Transactional Interpretation of Quantum Mechanics", Reviews of Modern Physics 58, 647 (1986); or: http://faculty.washington.edu/jcramer/TI/tiqm_1986.pdf

[Hobson] Art Hobson, Realist Analysis of Six Controversial Quantum Issues, https://arxiv.org/pdf/1901.02088.pdf

[Art H] Art Hobson "There are no particles, there are only fields" American Journal of Physics 81, 211 (2013) or https://arxiv.org/pdf/1204.4616.pdf

[Bell] John Bell, Against Measurement, Physics World, August 1990 p 33. https://m.tau.ac.il/~quantum/Vaidman/IQM/BellAM.pdf

[Nuclear] Nuclear Structure, https://en.wikipedia.org/wiki/Nuclear_structure

[**Herzog**_94] T.J. Herzog, J.G. Rarity, H. Weinfurter, and A. Zeilinger, "Frustrated Two-Photon Creation via Interference," PRL 72, #5 31 Jan 1994. Article in: https://www.univie.ac.at/qfp/publications3/pdffiles/1994-04.pdf And more discussion found in: Thomas J. Herzog, Paul G. Kwiat, Harald Weinfurter, and Anton Zeilinger, "Complementarity and the Quantum Eraser," Physical Review Letters, Vol. 75, No. 17, 23 October, 1995, Institut fur Experimentalphysik, Universitat Innsbruck
Paper Also at: http://isis.ku.dk/kurser/blob.aspx?feltid=17775 (ISIS, Copenhagen University). [This was an amazing experiment]

[FQXI] Dave Peterson "Physics Lives in Form Heaven," April, 2015, Posted at **FQXI Foundational Questions Institute** Essay Contest http://**fqxi.org**/community/forum/category/2351 (9 pages). "And, "What Fundamental Should Mean," **FQXi** Essay 12/14/17 https://fqxi.org/community/forum/category/31426 10 pages.

[Learning] Dave Peterson, "Learning Quantum Mechanics and Relativity," May 2015, (26 pages) www.sackett.net/LearningQuantumMechanics.pdf (26 pages)

[Keynes] John Maynard Keynes: Newton, the Man 1946 http://www-history.mcs.st-and.ac.uk/Extras/Keynes_Newton.html

[Found] Margenau, H. & Yourgrau, W., “Without Apologies,” Foundations of Physics, VoL 1, No. 1, 1970,
https://page-one.springer.com/pdf/preview/10.1007/BF00708649
https://link.springer.com/journal/volumesAndIssues/10701

[Jackson] David J. Jackson, “Octonians in Particle Physics through Structures of Generalized Proper Time,” ArXiv:1909.05014 9/2/19.

[Furey] C. Furey, Standard model physics from an algebra?
https://arxiv.org/pdf/1611.09182.pdf Waterloo, Thesis on Octonians and Dixon algebra D = R⊗ C⊗ H⊗ O .

[Frankel] Theodore Frankel, The **Geometry of Physics**, An Introduction, 2nd Ed. Cambridge, 1997, 2004.

[Jones] Mike Jones, Quantum Gravity and Machs Principle [Book], essays from 1967 to 2019. On the Web at:
http://cires.colorado.edu/~mjones/physics/Quantum%20Gravity%20and%20Machs%20Principle%202019.pdf

[Meg] E. Megidish et. Al., “Entanglement between photons that have never coexisted,” https://arxiv.org/abs/1209.4191, 2012.

[Riess] “Tensions between the Early and the Late Universe,” Kavli workshop, July 2019, ArXiv:1907.10625

[Kerr] Matt Visser, “The Kerr spacetime: A brief introduction.”
https://arxiv.org/pdf/0706.0622.pdf

[Afriat] Alexander Afriat, Quantum waves in configuration space, (London thesis 1999) http://etheses.lse.ac.uk/2085/1/U551423.pdf 143 pages.

[Kastner] https://www.informationphilosopher.com/solutions/scientists/kastner/
Ruth Kastner, Books: Understanding Our Unseen Reality: Solving Quantum Riddles 2015. And, The Transactional Interpretation of Quantum Mechanics, 2013.

[Sciama] D. W. Sciama, “On the Origin of Inertia,” Monthly Notices of the Royal Astronomical Society, Vol. 113, p.34, or see:
http://adsabs.harvard.edu/full/1953MNRAS.113...34S

[Schiff] Schiff L I 1960 “On Experimental Tests of the General Theory of Relativity,” Am. J. Phys. 28 340

[Liddle] Andrew Liddle, An Introduction to Modern Cosmology, 2nd Ed., Wiley, 2007.

[Pettini] M. Pettini: Introduction to Cosmology — Lecture 2 NEWTONIAN COSMOLOGY https://www.ast.cam.ac.uk/~pettini/Intro%20Cosmology/Lecture02.pdf

[CI] Stanford Encyclopedia of Philosophy, “Copenhagen Interpretation of Quantum Mechanics,” 2002, 2014. https://plato.stanford.edu/entries/qm-copenhagen/

[Watkins] Thayer Watkins, “Evidence for Alpha Particle Substructures in Nuclei,”
http://www.sjsu.edu/faculty/watkins/alphastruct

[Atiyah] Memorial tribute to Michael Atiyah, Notices of the American Mathematical Society, Dec. 2019, p 1841.
{“Michael felt that the four division algebras – real and complex numbers, quaternions, and octonians – provided essentially the only mathematically natural way to account for the number of fundamental forces (four) or the number of generations (three) in the Standard Model.”}

Note: slash partial $\not\partial$ is called the “Dirac Operator” [insert field eq o(O,/), o=overstrike, $\not{o}$]

LEARNING QUANTUM MECHANICS AND RELATIVITY

DAVE PETERSON

Abstract. Students of quantum mechanics and relativity encounter material that is not only mathematically difficult but also conceptually incredulous. They ask, "Isn't there a way to make the mathematics and its interpretation more transparent?" and also, "Is this the way the world really is?" Standard texts on non-relativistic quantum mechanics (QM) generally focus on the abstract mathematical machinery for solving problems and usually minimize or avoid attempts at intuitive understanding, basic underlying physics, and the existence of many differing interpretations of the mathematics. Also, different texts present quite different sets of fundamental but abstract postulates for a coherent system from which to calculate probabilities of experimental outcomes. The goal of this paper is to balance out that strangeness and abstraction by providing an intuitive understanding of some of the key parts of this machinery. We wish to motivate and simplify so that the mathematics isn't quite so formidable. After this introduction, standard texts may be studied in the usual way.

Special and general relativity also have their abstractness and opaqueness. Why, for example, do their metrics have differences in sign between space and time parts? And, when dealing with weak gravitational fields, is it still necessary to know the language of tensor calculus or advanced differential geometry?

Most of the special heuristic tidbits discussed below are not well known. Why that should be is largely a mystery to me.

1. Introduction

The mathematical theory of quantum mechanics is highly successful and has flawlessly passed nearly ninety years of careful experimental tests. College textbooks on quantum mechanics generally do a good job of providing adequate coverage of topic material so that students have a conventional common mathematical machinery for solving relevant physics problems. But this is often done in a sparse fashion which presents abstract postulates and rules without sufficient motivation or physical clarity. They don't say why we do things in this conventional semi-Copenhagen way, how much linear algebra one should have first and why, what's really going on, where's the physics beneath the abstract mathematics? They presume that the machinery will make some sense (or at least familiarity) after solving a series of problems. But basic postulates and math are counterintuitive and are given "out of the blue;" and application is done from abstract generals to particular examples. It is fairly easy to claim that in quantum mechanics, heuristics are poor. But there is also a reason for this: physicists do not agree about the possibilities for any underlying reality.

Date: July 8, 2014.

email: davepeterson137@gmail.com. Paper updated to 4/17/15, May 4, 2015.

Some even claim that there is no underlying reality. And there is no general agreement on the interpretation of the mathematics of quantum mechanics. As an example, one reason that Niels Bohr was so difficult to understand was that he was very careful to always avoid any mention of possible underlying mechanisms. So, if you are one of those people who ask, "What is really going on?," you may find few answers.

Every text on quantum mechanics provides a list of postulates from which basic structure can be developed and problems solved for the probabilities of experimental outcomes. Postulates are often stated with numbers (like [P1]) with orderings and main choices that vary from text to text. Here, I will assign numbers to primary postulates, but the choice is fairly arbitrary. Postulate 1 [P1] of quantum mechanics is always about the existence of a complex state function, $\Psi(x,t)$ (or "ket" $|\Psi\rangle$ in Dirac notation), describing any physically-realizable state of of a system and claiming that it contains all accessible physical information about that system [1] . Sometimes, this is accompanied by "The principle of Superposition" being added that for any physically realizable states, other states can be formed by linear superposition with complex valued coefficients. These complex coefficients stress the importance of the relative phases of the components being added [2]. The other postulates appear with different numbers from text to text so that their names (if provided at all) are more important than their numbers.

The ordering I will use here for the primary postulates of quantum mechanics are: again [P1] for the existence complex state functions, [P2] is about operators corresponding to observables, [P3] is the Schrödinger equation, [P4] says that measurement is a projection for resulting eigenvalues, [P5] is the Born Rule for outcome probabilities, and [P6] says that states of a composite system are tensor products of component states.

In my view, the strangest and most important postulate is the **"Born Rule" [P5]** implying a probabilistic interpretation for the wavefunction [3]. In one simple example, $|\Psi|^2$ is the probability that the system will have given coordinates at time t; and this in turn means that Ψ is a strange and new concept called a **"probability amplitude"** (at least in the usual Copenhagen interpretation of quantum mechanics). It also means that the sub-quantum world, if it indeed "exists" at all, lives in something like "the square root of reality." And that often makes it very different from any concept in classical physics. In

[1] Schrödinger initially intended his Ψ to correspond to a real wave, but Born's probability wave quickly prevailed instead. Students of Schrödinger wrote a poem: "Erwin with his psi can do, Calculations quite a few. But one thing has not been seen: Just what does ψ really mean? [Remembered by Felix Bloch].

[2] i.e., a complex coefficient can be written in polar form, $c_1 = a_1 e^{i\phi_1}$, where angle ϕ represents the relative phase difference of functions with respect to neighbors in the sum. They are all locked-in or entrained together with these fixed relative phases. An imaginary coefficient implies phase 90^o *or* $\pi/2$ radians, i.e., $i = e^{i\pi/2}$

[3] Max Born stated this conclusion as a footnote in a 1926 paper on particle collisions. Yes, I know that Schrödinger said that entanglement was the strangest and most distinguishing concept; but I would call it just a close second.

particular, it makes the use of complex and hypercomplex numbers seem to be a necessity (e.g., quaternions and Dirac matrices) [4]. It should be a goal of new research into the foundations of quantum mechanics to derive the Born Rule instead of simply postulating it.

One factor making the quantum postulates inconsistent from text to text is that they are a mishmash of main postulates (that are logically fundamental), secondary postulates (derivable from main postulates), and other "mere consequences" of postulates that happen to be well known (like the "uncertainty principle" which is just a derived principle rather than being fundamental) [1]. Inconsistent stress makes it unclear what is most important. Unlike the principles underlying relativity (see later section), the principles of quantum mechanics are exclusively in the language of abstract mathematics whose physical meanings are unclear. The corresponding physical principles are in dispute.

Complex numbers in quantum mechanics: Complex numbers appear almost everywhere in quantum mechanics and greatly facilitate calculations. Quantum formulations depend on the use of complex numbers in all textbooks. So, if you want to learn quantum mechanics, you have no present choice but to accept and use complex numbers. In disciplines such as electrical engineering, complex numbers are a great convenience in calculations; but final answers just use the real part. A strong majority opinion among physicists is that complex numbers are instead essential and intrinsic in quantum mechanics. In the discussions below, the initial choice of describing waves in complex polar form, $ce^{i\varphi}$, leads to the use of complex amplitudes and then complex operators. Adding waves of different shapes or frequencies means caring about the phase relationships between waves, and complex numbers do that well. The resulting mathematical system is highly dovetailed, self-consistent, and tremendously successful. There are still many dissenters who wish to structure quantum mathematics differently (such as using 2×2 matrices in place of complex numbers). But their attempted constructions generally increase computational difficulty and reduce economy of the mathematics. Ultimately, the dogma of complex numbers being intrinsic depends on finding a good interpretation of the quantum world (e.g., is the wave-function real in some sense (ontology) versus having it rather reflect "our knowledge" of a system (epistemology).

Here is an **Outline** of key points addressed in the sections that follow:

[4]For example, electron spin has a "Hilbert Space" of only two base vectors, $|up\rangle$ and $|down\rangle$ for a spin projection in say the "z" direction. But after a test with a Stern-Gerlach magnet, future spins can be measured in an x or y direction too. The two z-bases cover both of those cases (very un-real-vector-like behavior because spin is hyper-complex). If we let $a = 1/\sqrt{2} \simeq 0.707$, then x-spin right $= |\rightarrow\rangle = a|\uparrow\rangle + a|\downarrow\rangle$ and spin y or spin down into the paper $|\odot\rangle = a|\uparrow\rangle + ia|\downarrow\rangle -$ funny superpositions of up/down base states. And for y, a complex coefficient is really required. Also, the y-spin operator (Pauli Matrix) is complex: $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$. This can be thought of as i times a quaternion, q_y.

- The simple plane wave is used to intuit differential **operators** on eigenfunctions to give eigenvalues. Conservation of energy is then written with operators to form the Schrödinger Equation of quantum wave-mechanics and the unitary evolution of state functions.
- Simple examples are given of superposition. The Born Rule explains chemical bonding by an extra enhancement of electron density between two nuclei.
- The uncertainty principle is derived as a consequence of primary postulates and shown in two forms.
- Further mention is made of the primary postulates of quantum mechanics: [P1] complex state function, [P2] corresponding operators, [P3] Schrödinger equation, [P4] projection for resulting eigenvalues, the Born Rule [P5], and tensor products for composite systems [P6].
- The elementary non-relativistic Lagrangian, $L = T - V$, is derived simply by **counting waves along a path** (making the simplest Feynman path integral easy to understand). And later, it is derived from the "principle of maximum proper time" along a path.
- Time dilation, $\Delta t = \gamma \Delta \tau$, is derived simply from the Lorentz metric designed to give zero interval when two events are connected by light.
- Length contraction is derived simply from the invariance of the metric (with a standard derivation shown in the appendix).
- The relativistic Lagrangian, $L = -m_o c^2/\gamma - V$, is derived simply by counting waves along a path.
- First order general relativity is explained simply by special relativity combined with the principle of equivalence.

Many of the items mentioned here and in the following are not well known and are generally hard to find in the literature. I had the joy of discovering them largely by myself. But I presume that they are familiar to "those who know well."

2. Background

I believe that the most important concepts in quantum mechanics should begin with stating:

$$(1) \qquad [P0] \qquad \mathbf{p} = \mathbf{h}/\lambda = \hbar\mathbf{k} \qquad \textit{and} \qquad \mathbf{E} = \mathbf{h}\nu = \hbar\omega \ ;$$

and these equations apply to both light waves and to matter waves [5]. I would call this Postulate Zero [P0], and it is a statement in the language of physics. No textbook incorporates this as a postulate largely because it suggests that waves have a reality; and that goes against the majority beliefs of the mid-1900's. But, if we wish for an intuitive understanding of QM, we must start with the belief that these waves are at least"real" in some sense, and that view is becoming increasingly more popular. A big question is "waves of what?" The orthodox answer is "waves of probability amplitude" − but that certainly

[5] Here, $\hbar = h/2\pi$, $k = 2\pi/\lambda$, $\omega = 2\pi\nu$ and non-relativistic momentum $p = mv$.

doesn't resonate in our intuitions. For light photons, the waves appear to just be vector electromagnetic or vector-potential waves. This is especially apparent when single photons refract through glass in the same way that classical electrical waves do [6].

The key equation $E = h\nu$ originated with Planck's 1900 paper on black body radiation for what were later called photons. The quantum idea was that if a frequency ν was present, then it was only capable of delivering a "quantum" of energy, $\Delta E = h\nu$ to an absorber. Then, in 1924, de Broglie used Einstein's relativity theory to claim that massive particles also obey this rule and that total energy determines a fundamental clock rate for electrons. He also claimed that an effective wavelength exists for moving particles, $\lambda = h/p$, extended from Einstein's idea that light quanta also possessed momentum [7]. We later said that massive particles with momentum and energy have associated de Broglie matter waves with a wavelength and frequency.

That electrons diffract from crystals just like x-rays was first shown by Davisson and Germer in 1927. It was then shown that matter wave diffraction also occurs for neutral atoms and now even large molecules like buckyballs (C_{60}). Instead of being physically "real" waves, the scalar "matter" wave might be understood as representing an information "code" where wave concentration in space can inform about momentum and wave peak density in time tells energy. A wave also has a phase-velocity given by $v_\phi = \lambda\nu = (2\pi\nu)(\lambda/2\pi) = \omega/k$. So, a wave moving to the right would be given by x-coordinate $x = v_\phi t$, *or* $kx = \omega t$. We pick a point or phase on the wave and follow its motion. So, if $y = a\cos\phi = a\cos(kx - \omega t)$ has say $\phi = 0$ (and peak $y = a$), we can follow its motion to the right.

What are the most important equations in mathematics? The Pythagorean theorem might be one answer (although it only applies when space is flat and not, say, on the curved surface of the Earth). But competing with that answer might be Euler's formula, $e^{i\pi} = -1$ (relating the number $e \sim 2.718$ and named "e" after Euler, and pi, and the 'irrational' number i). If it is that important, then almost everyone should know it. And it is a special case of $e^{i\theta} = \cos(\theta) + i\sin(\theta)$, which gives Euler's formula when the phase is pi (180 degrees). Rather than having just real waves, we prefer to generalize to complex waves using the exponential with base e. This is not just for the usual calculus convenience; it is widely believed that complex numbers are intrinsic in quantum mechanics. So, for matter waves, we write: $y = a\, e^{i\phi(x,t)} = ae^{i(kx-\omega t)} = \mathbf{ae^{-i(\omega t - kx)}} = \mathbf{ae^{(-i/\hbar)(Et-px)}}$.

In quantum mechanics, we like to label our wave functions with amplitudes and phases by the symbol, "psi," Ψ. The equation above is for an infinitely long "plane" wave (over all x and all t). If it also represents the motion of a so-called "particle," we might want

[6] but we then have to ask ourselves which came first: classical EM waves down to the quantum level or intrinsically quantum EM ideas seeming classical due to large numbers of boson photons. That answer seems to be quantum first, and then build up from there.

[7] First noticed by Stark in 1909 as $p = h\nu/c$ and then finally and formally by Einstein in 1916.

to restrict its domain better and "localize" it. This usually involves some superposition of other wavenumbers, k (and discussed under the topic of Fourier analysis). A gaussian shape in space for the waveform associated with a particle would result from a bell-shaped profile of wave numbers, k. But, for the present, we will just look at the oversimplified plane wave.

Once we have wave phases in an exponent, we will wish to be able to pull down the values of E and p (their eigenvalues) from the expression for the wave. Obviously, this can be done by creating and applying "operators" $\hat{E}$ and $\hat{p}$ so that $\hat{E}\psi = E\psi$ and $\hat{p}\psi = p\psi$; we just make operators that work that way.

3. Simple Plane Waves:

Using a plane-wave traveling wave train as the most elementary heuristic example, we have a choice of expressing it as a wave in its own terms or in terms of energy and momentum as parameters a measurement observation might prefer.

$$\psi(x,t) = Ae^{-i(\omega t - kx)} \;\rightarrow\; \psi(x,t) = Ae^{-i(E_o t - p_o x)/\hbar} \tag{2}$$

Since $E = \hbar\omega = h\nu$, and $p = h/\lambda = \hbar k$, these equations are equivalent. In this equation, we have single constant values for the wavenumber, k, and the angular frequency, ω. It might be that the left equation happens to be the one preferred by Nature for a wavefunction in the spacetime between an emitter and detector and that the Planck constant, $\hbar$, might only enter when a (classical) detector 'collapses' the wavefunction to make use of its particle energy or momentum [8]. Perhaps the simple wave is everywhere a carrier of quantum information without physical actualization; and the density of wave peaks in space and in time represents information as a **'code'** about what might actually be detected as a physical particle.

Nature can also use this code to deduce a particle's rest mass, see for example equations (15) and (16) for m_o and ω_o later on. The 'particle' itself is only a deduction by the measuring apparatus and likely doesn't exist physically in the wavefunction. The amplitude of the wavefunction can disperse and weaken over time and distance and still carry the information ultimately used. Note that the units of h are $[h] = joules \cdot sec = J/hertz = [momentum]/wavenumber = [action]$. Each vibration per second contributes a unit of energy; each packed wavelength adds momentum.

Operators: If one begins with $\psi = \psi(x,t)$ as in equation (2), we then wish to retrieve the energy and momentum it contains in the exponential. Obviously, derivatives will pull these out. That is, creating an **operator** denoted as $\underline{\hat{\mathbf{p}} = -\mathbf{i}\hbar\,\partial/\partial\mathbf{x}}$ (or $\hat{p} = -i\hbar\nabla$ in 3-dimensions) gives us a so-called "eigenvalue" equation $\hat{p}\psi = p_o\psi$. That is, the momentum operator on the wave function yields a constant times the wave function.

[8]It seems to me that should require some sort of sub-quantum network transaction or hand-shaking agreement between source and absorber. But due to a general avoidance of discussion about mechanisms, that is a minority opinion.

And using operator $\underline{\hat{\mathbf{E}} = (\mathbf{i}\hbar)\,\partial/\partial\mathbf{t}}$ gives us $\hat{E}\psi = E_o\psi$ [9]. This is a first example of another postulate in QM that might be called the "correspondence principle" [10] :

[P2] To every physical observable, there corresponds a linear operator (and we call it "Hermitian" or self-adjoint if it always results in a real value for the observable [11]). It is said that a difficulty in understanding QM is that instead of momentum being a deterministic variable, it now IS an operator operating on a wave function. I think what that means is that the operation of the operator interprets and enables activation of the underlying code contained in the wave function (the density of wave peaks over distance).

So far, there is nothing proprietary about having linear operators for quantum mechanics. We can also have them for classical waves too. In equation (2), for example, we could create operators that pull down the value of the angular frequency, omega, or the wave-number, k, by using $\hat{\omega} = i\,\partial/\partial t$ and $\hat{k} = -i\,\partial/\partial x$. Then, $\hat{\omega}\psi = \omega\psi$ and $\hat{k}\psi = k\psi$. Again, this is an example of a postulate for what is called a "linear eigenvalue equation" associated with each linear operator [2]. Ψ is called an eigenfunction of the operator, and the real constant is called the eigenvalue. It is also a postulate that [P4] "one or another of the eigenvalues is the only possible result of a precise measurement of the dynamical variable represented by" the linear operator [2].

Some texts consider **Schrödinger's equation** of 1926 ("SE," eqn.(3) below) as a primary postulate of QM, [P3]. Here, it is simply obtained by writing out conservation of total energy for a single particle in terms of these new operators, $\hat{p}$ and $\hat{E}$, on a wave function. Since kinetic energy $KE = mv^2/2 = p^2/2m$, the operator for KE will be $\hat{KE} = \frac{1}{2m}\hat{p}^2$; and potential energy $\hat{V} = V$. And these operate on the wave function, $\psi(x,t)$:

$$(3)\;\; KE + V = \frac{p^2}{2m} + V = E_{total} \;\rightarrow\; -\frac{\hbar^2}{2m}\nabla^2\psi(x,t) + V(x,t)\psi(x,t) = i\hbar\frac{\partial\psi(x,t)}{\partial t} \quad [P3].$$

It has always seemed to me that this simple approach is the best way to intuitively introduce the Schrödinger equation for the first time rather than just postulate the strange complex-looking Schrödinger equation and have it sprung onto a first time reader [12]. What the $\hat{p}$ operator does is look at the density of wave peaks in space, and the $\hat{E}$ operator looks

[9] And for angular momentum, L, one considers change of phase around a circular phi direction, $\partial/\partial\phi$, or in more generally in 3D by $-i\hbar\, r \times \nabla$.

[10] But, the term "correspondence principle," is also used to state that the predictions of QM reduce to those of classical mechanics in the limit where a system approaches large quantum numbers or higher energies.

[11] Unlike the complex quantum world, the classical world only desires real results.

[12] And Weinberg's text on QM does touch on this heuristic introduction [3]. His book is also one of the few to mention interpretations (section 3.7) − but only for Copenhagen, Many Worlds, older hidden variables, and Decoherent Histories. He adds: "My own conclusion (not universally shared) is that today there is no interpretation of quantum mechanics that does not have serious flaws, and that we ought to take seriously the possibility of finding some more satisfactory other theory, to which quantum mechanics is merely a good approximation."

at the density of peaks in time. I've always thought that Nature must also do this by phase comparisons over small space-time regions. So, from a wave, Nature can deduce E and p.

The observable operator interpreted to mean energy (such as $KE+V$), is a distinguished observable called the 'Hamiltonian,' $\hat{H}$. So the Schrödinger equation can also be written as:

$$\hat{H}\psi(x,t) = i\hbar\frac{\partial}{\partial t}\psi(x,t), \quad or \quad \frac{\partial \psi}{\partial t} = \frac{-i}{\hbar}\hat{H}\psi \quad so, \quad \psi = \psi_o e^{-iHt/\hbar} = U(t)\psi_o(x,0) \tag{4}$$

where U(t) is a unitary time evolution operator that can be written as an exponential. So, we can use the Hamiltonian to give the time evolution of the wavefunction, $\psi(x,t)$. In some texts, this **unitary evolution with time, U(t)**, is given as a primary postulate **also [P3′]**, and the SE follows from it. Hamiltonian energy isn't always $KE+V$, there are other forms too. For example, a particle with a magnetic moment in a magnetic field may have the form: $\hat{H} = \mu B \cdot \sigma$ (where σ refers to Pauli matrices). And when electromagnetic fields are present, $\vec{p}$ becomes $\vec{p} - e\vec{A}$. But, for any Hamiltonian energy, $\psi(x,t)$ evolves continuously and deterministically into the future, until the point where final measurement occurs. Then the wave function collapses, and determinism is lost. In other words, "a great miracle occurs," and nobody really knows how.

Unlike the simple example above, in traditional classes the Schrödinger equation is simply presented as (an initially strange) founding postulate of non-relativistic quantum mechanics. Its solutions include tunneling, complex atoms, and s-orbitals which no longer resemble anything like plane waves. For example, just try a solution resembling an exponentially decaying radial profile: $\psi_1 = Ae^{-b\,r}$ and plug that into the SE with an atomic nucleus central potential $V = -Ze^2/4\pi\epsilon_o r$ and use $\nabla^2\psi = r^{-2}\partial/\partial r(r^2\partial\psi/\partial r)$. And then solve for the actual coefficients A and b. The result is the normalized [13] 1S atomic orbital:

$$\psi_1(r) = \frac{1}{\sqrt{\pi}}\left(\frac{Z}{a_o}\right)^{3/2} e^{-Zr/a_o}, \tag{5}$$

where $a_o = 4\pi\epsilon_o\hbar^2/me^2$ is the first Bohr orbit $\simeq 0.53\mathring{A}$, and Z is the proton number [14]. And then there is also multiplication by a time varying factor with a frequency given by $\nu = E/h$. $\psi_1(r,t)$ is like a tent shape that is up and then becomes inverted down and then back to up again − but in 3D. This profile is like nothing experienced in the classical world, and there is nothing orbiting in the orbital. So, even though the Schrödinger equation makes simple intuitive sense for plane waves, its application goes well beyond that.

[13] In quantum mechanics, normalizing means finding a front end coefficient such that the integral of $\psi^*\psi$ over all space gives 1 = 100% total probability.

[14] This is the innermost atomic orbital that Bohr missed in his early theory where electrons were standing waves about a nucleus. Actually, he only quantized the orbital angular momentum without yet realizing that it could represent de Broglie waves. And he started with angular momentum 1, 2, and up; while the 1s state has orbital angular momentum $L = 0$.

So, introducing it as a postulate makes some sense [15].

This particular atomic 1S orbital can be superimposed with other orbital functions and still be an appropriate combined wave function. An example is the single electron shared by two hydrogen nuclei, A and B for the simplest molecule ${H_2}^+$, with binding wave function $\Psi_+ = a(1S_A + 1S_B)$. The "binding" itself is due to the overlap of these two spherical functions followed by squaring (the special enhancement of electron density in-between the two protons due to application of the Born Rule). Another is carbon's four-valence-electron hybrid orbital [Linus Pauling, 1931]: $\Psi_1 = 0.5(2s + 2p_x + 2p_y + 2p_z)$ where the p-orbitals have angular momentum $L = 1\hbar$. In both cases, the coefficient signs (+ in this case) are very important because they represent the coordinated phases of the superpositions. The result for carbon is the formation of a lobe of enhanced electron density sticking out from the carbon atom in the direction $\hat{i} + \hat{j} + \hat{k}$. For all four electrons, we get four tetrahedral spaced lobes ready for bonding (e.g., like for methane, CH_4). In carbon, rather than have all those individual orbitals vibrate separately, it makes sense for them to get entrained together (entangled or hybridized) so that they have more aspects of constructive interferences.

Some say that the uncertainty Principle is a key property of QM, and that is occasionally introduced as a postulate too. But it is actually just a derivation from other more key postulates. In the oversimplified case of a plane wave, there is no localization of any presumed particle. Localization can be expressed with a wave-packet which can be created from a Fourier distribution of plane waves. If the wave-packet has a spatial width (say the standard deviation for a Gaussian packet), then the uncertainty principle applies in either form for the widths of x versus p or for x versus wavenumber, k (i.e., quantum mechanics not required).

That is, somehow, Nature effectively can perform the equivalent of Fourier Transforms (going from waveform in space or time to wavenumber or frequency in space or time). It is not clear how it does this, but it explains the Heisenberg Uncertainty Principle. That is, let wave-packet shape have an associated Gaussian probability envelope such that its probability density, $P_x = dP/dx$, is described by:

$$P_x \propto e^{-x^2/2{\sigma_x}^2},\ so,\ \psi(x) = \sqrt{P_x} \propto e^{-x^2/4{\sigma_x}^2} \tag{6}$$

The symbol sigma refers to "standard deviation" or square-root of variance in statistics. The Fourier Transform (FT) of a Gaussian is itself a Gaussian so that the momentum wavefunction $\phi(k) = \sqrt{P_k} \propto \exp(-k^2/4{\sigma_k}^2)$. Since $\exp(-a^2x^2) \longleftrightarrow \exp(-k^2/4a^2)$ is a transform-pair where $a^2 = 1/4{\sigma_x}^2$, we have:

[15]There is a Fourier Transform from the 1S exponential decay wave, but it is a 3-D spherically radial transform not easily associated with plane waves. The form in momentum space is $\phi(\vec{p}) \propto p/(p^2+1)$ [4]. One has to integrate $\int \psi(r) exp(-ip\cdot r)dr$.

$$\frac{k^2}{4(1/4{\sigma_x}^2)} = k^2{\sigma_x}^2 = \frac{k^2}{4{\sigma_k}^2}, \Rightarrow \ \sigma_x\sigma_k = \frac{1}{2}, \Rightarrow \ \sigma_x\sigma_p = \frac{\hbar}{2}, \ \ or, \ \Delta x\Delta p = \frac{\hbar}{2}. \tag{7}$$

The case of Gaussian envelopes is optimal and gives equality. Any other waveform envelope profile will give $\Delta x\Delta p > \hbar/2$. A distribution of momenta in a wave packet will cause spreading of the spatial width of the wave packet over time

In the case of just waves without momentum being considered, Fourier transform theory says, $\Delta x\Delta k > 1/2$. That is, an uncertainty principle applies to waves by themselves without any mention of Planck's constant, $\hbar$. In electrical engineering, "It is well known that the bandwidth-duration product of a signal cannot be less than a certain minimum value" [5]. That is, $\Delta t\,\Delta freq \geq 1/4\pi$ or $\Delta t\,\Delta\omega \geq 1/2$. So, if a special class of electrical engineers had existed in 1927, there wouldn't have been so much mystery about these uncertainty principles.

Another important concept in quantum mechanics is the use of the "commutator bracket" of two linear operators: $[\hat{A},\hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}$. Most of the time in classical mechanics, the commutator will be zero. But, in QM, $[\hat{x},\hat{p}_x]\psi = -i\hbar(x\partial_x\psi - \partial_x(x\psi)\ ) = i\hbar\psi$. This is used in a general form for uncertainty relations, $\Delta A\ \Delta B \geq \langle [A,B]\rangle/2i$ so that $\Delta x\Delta p \geq [\hat{x},\hat{p}]/2i = i\hbar/2i = \hbar/2$. But again, neither "commutators" nor "the uncertainty principle" are unique to quantum mechanics, they also appear in usual classical physics [6]. For example, for classical waves, $[t,\partial_t] = -\mathcal{I}$. What is unique to QM is the appearance of the value $\hbar$, the concept of "entanglement," the existence of probability amplitudes, and the phenomenon of "collapse" of the wave-function and the apparent reification of particle behavior.

4. The Postulates

The first postulate of quantum mechanics is sometimes stated more elaborately as:
$[P1]$ For every system, there is a corresponding **Hilbert space**, $\mathcal{H}$ [16]; and a state of the system is a unit ray in the Hilbert space.
A student has to understand this statement but might also ask why it is written in this

[16] In 1932, von Neumann decided to include a collection of states into a "Hilbert Space" from a publication in 1924 by Courant and Hilbert (for pure mathematics purposes). A Hilbert space is an abstract vector space having an inner product. The simplest example is the ordinary real Euclidean vectors with unit vector basis $\{i,\ j,\ k\}$ and the familiar dot product $\vec{A}\cdot\vec{B} = |A|\ |B|\cos\theta$. In quantum mechanics, we can add that it is a "function vector space" [such as spherical harmonics, Hermite polynomials, or Legendre polynomials (1782) on the interval $-1 \leq x \leq 1$] and referred to under the heading of "Lebesgue spaces, L^2 " of square integrable functions. But quantum mechanics allows for complex coefficients. This applied mainly to Schrödinger's Wave Mechanics. Heisenberg's Matrix Mechanics came a little earlier in 1926 and was formulated with potentially infinite square matrices with a Hilbert Space of sequences of complex numbers: "little ℓ_2" spaces. The two formulations are mathematically equivalent. Note that Hilbert space can be, and often is, composed of an infinite number of bases. So, a vector can be a sum of an infinite number of components.

initially opaque language. Hilbert space includes linear vector spaces, so the usual mathematics from a course in "Linear Algebra" is automatically implied. This, of course, includes the superposition principle (adding vectors together still gives a vector). The "vectors" in this case are more commonly functions that can be added and subtracted in the same way as vectors (except that phase is also important).

Two State Superpositions: One of the simplest examples of postulate [P1] is the polarization states of the photon (e.g., see Feynman Lectures Vol. III [15]). For the case of a photon traveling in the z-direction, the Hilbert Space of this single photon system can simply consist of two basis vectors called $|x\rangle$ and $|y\rangle$. That means that the electric field vector is perpendicular to the direction of motion and can be in the "horizontal" x or "vertical" y directions. Any other direction (like 45^o) is a real superposition of these base states. A horizontal polarizer will not pass the $|y\rangle$ state. But, we also like to say that a photon carries spin; and this can be written as right or left circularly polarized states by a complex superposition of the base states:

$$|RHC\rangle = |R\rangle = \frac{1}{\sqrt{2}}(|x\rangle + i|y\rangle), \qquad |LHC\rangle = |L\rangle = \frac{1}{\sqrt{2}}(|x\rangle - i|y\rangle), \tag{8}$$

where the $1/\sqrt{2}$ coefficients "normalize" the states (i.e., $R^*R = 100\%$ and $L^*L = 1$). The imaginary coefficient i says that the addition of the y sine wave is 90^o out of phase with the x sine wave (so the y-wave is cosine). When that happens, the tip of the electric field vector, $\vec{E}$, rotates about in a helix and carries angular momentum.

Now, equation (8) can be inverted to solve for x and y in terms of R and L. That means that R and L could also be considered as the bases for the Hilbert space. So, which is more "real?" Both selections are equally valid with utility varying with the nature of the experiment observing the photons (e.g., polarizers or quarter-wave-plate/Nichol-prism combination, etc.). Single photons can be either circularly (or elliptically) polarized or linearly polarized. So, how can photons have spin $S = \pm 1\hbar$ and also be linearly polarized? The answer is that a linearly polarized photon (spin zero) can be considered as a superposition of both forward and reverse spin (RHC and LHC) at the same time. That is OK in quantum mechanics.

And, as if that wasn't counter-intuitive enough, we can also have macroscopic cases of persistent currents in superconducting ring loops that can exist in a superposition of both clockwise and anti-clockwise directions of current flow at the same time.

"State" is a key word in quantum mechanics. In Schrödinger "wave mechanics" it may also be called a "wave function." A traveling (time dependent) state is a mathematical expression for a matter-wave that represents an appropriate relation or transition between a source and a detector and possibly what's in-between. It has to go through mathematical processing's before it can be said to have any classically understood "reality" (unitary evolution, "reduction," Born rule "squaring,"...). There are also time-independent states such as the hydrogen atom orbitals, and these can be considered as "standing-waves." Exactly what a state means and how "real" it is in itself has been a source of continuing discussion and ongoing contention. The mathematics has always worked perfectly, but what a state represents to us is somewhat opaque. In a sense, it tells all of the possible outcomes from a measurement. Examples include the interference output from slits in two-slit diffraction, a moving electron, electron spin, atoms, and molecules. Measured values of an experiment are called eigenvalues which are intrinsically classical and real and are not properties of quantum objects which are complex [7].

The most desirable background for studying quantum mechanics is a mathematical knowledge of linear algebra. In older days, students simply picked this up during the learning of quantum mechanics. This has the advantage that only a portion of linear algebra is needed, and in physics that portion is in the desired notation and application (our vectors are in Hilbert space, a term barely mentioned in math books). One nice source for learning this is in Griffiths [18]. After a course of study, one should be able to easily say things like, "two unequal eigenvalues of a Hermitian operator have orthogonal corresponding eigenvectors." In most cases of interest, a state function $|\psi\rangle$ will have a variety of eigenvalues and eigenfunctions $-$ not just one as for the simple plane wave. Then, a better statement of the Born Rule [P5] is: given that a system is prepared in a state $|\psi\rangle$, the probability of seeing a measured system in an eigenstate $|a\rangle$ for an observable $\hat{A}$ is given by $P_a = |\langle a|\psi\rangle|^2$. We have to know about inner products, linear functionals and the dual space, operators, projectors, subspaces, orthonormal bases, matrices, diagonalization of Hermitian operators, traces, density matrices, probability theories, tensor products, and much more [18]

A very key (and very confusing) term in quantum mechanics is "measurement." One view is that it is a projection operating on the wave function and always causing the system to jump into being an eigenstate of whatever dynamical variable is being measured. The measured result is the eigenvalue of that eigenstate. A measurement actualizes values for the state. A standard view is that a state, ψ, **collapses** its wave-function in the act of measurement. A wave-function may be spread over kilometers (or possibly even light-years) but then has to suddenly everywhere collapse into a point for measurement detection: "The electron or photon ended up Here!" Possible mechanisms for doing this are presently unknown, and there are many conceptual difficulties (the "Measurement Problem"). If we are expecting an explanation to connect the measurement outcome to some property of a particle before the time of measurement, the problem might be in the word "before" (presumptions about the nature of quantum information and time). And the term collapse might be replaced by other suitable conditional probabilities in consistent theories [18].

Another way of stating measurements [P4] is the "von Neumann Postulate: If a measurement of the observable A yields some value a_i, the wave function of the system just after measurement is the corresponding eigenstate ψ_i [1]. This is another non-deterministic discontinuous collapse due to the act of observation by projection of a superposition to one of its terms.

In addition to the "main postulates" already mentioned ([P1] complex state function, we again also have [P2] corresponding operators, [P3] wave equation, [P4] projection for resulting eigenvalues, and the Born Rule [P5]) , one can derive "secondary postulates" [1]. These include superpositions, eigenfunction and eigenvalues, calculation of expectation values, expansion in eigenfunctions, and conservation of probability.

A further note on postulate [P6] Tensor Products: A primary postulate that sometimes goes unmentioned as primary is about "tensor products": the state of a composite system is in the direct product of the Hilbert spaces of its component systems: $S = S_A + S_B \implies \mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. This is important when discussing entanglements for two or more particles. For example, the state $|\Psi\rangle = (1/\sqrt{2})(\ |1\rangle_A \otimes |0\rangle_B - |0\rangle_A \otimes |1\rangle_B\ )$ is an entangled state.

Historically, Heisenberg's "Matrix Mechanics" came slightly earlier than Schrödinger's "Wave Mechanics," (1925 and 1926). But Schrödinger's mathematics was much easier to use and quickly gained popularity. Physicists were familiar with the language of differential equations, but few knew anything about matrices [17]. They were difficult to apply for most common problems and now find use in fewer applications (such as the harmonic oscillator). In 1926 to 1930, Dirac invented his more general "transformation theory" and could derive both the wave and matrix pictures from it (and operators now become the generators of transformations). Then in 1939, Dirac introduced his now common notation of "bra" and "ket" vectors, $\langle\varphi|$ and $|\psi\rangle$, with "inner product" then conveniently written as $\langle\varphi|\psi\rangle$. This is the analog of the usual "dot" product of vectors, $\vec{A}\cdot\vec{B} = |A||B|cos\theta$. But for continuous functions, it might look more like $\int \varphi^*\psi\ \ d(volume)$.

In addition to the wave and matrix formulations, there is also a "path integral" or "sum over histories" formulation from Richard Feynman (sum over all possible paths that a particle could take weighted by phases along each path). Feynman wrote a technical book on this [13], but he also discussed an elementary version in his much more popular book called "QED" [14]. From his formulation, he derived the Schrödinger equation. The relevant phases depend on "action," $A = \int Ldt$, where L was a slightly opaque function called a "Lagrangian."

Path Integrals and Least Action: Derive the Lagrangian $\mathbf{L} = \mathbf{T} - \mathbf{V}$ simply by counting waves along a path (making the simplest Feynman path integral easy to understand):

This is based on the "principle of least action" or "principle of stationary action" which dates at least back to 1662 for "Fermat's Principle" for light rays and to 1744 for massive

[17] Essentially, they had never been used by physicists since their discovery by Cayley in 1855 and were considered as "pure mathematics."

particles [Maupertuis and Euler]. The equations they used were:

$$\delta \int_{t_1}^{t_2} 2T(t)dt = 0 \qquad and \qquad \delta \int_{x_1}^{x_2} pdq = \delta A = 0. \tag{9}$$

where T is another symbol for kinetic energy, KE, "q" is a generalized coordinate symbol for distance (most of the time we could just use " x" instead), A is "action" (the integration over time), and trial paths are varied (symbolized by "change in" or delta . The game we play is to fix end points at time-1 and time-2 and then vary paths in-between until they satisfy requirements (called "calculus of variations" in mathematics with solutions given by the "Euler-Lagrange" equations). These concepts were then broadened by Lagrange [1760] and Hamilton [1835] where the integrand came to be called the "Lagrangian", $L = L(x, \dot{x}, t)$ which is often just $L = T - V$ (and here $\dot{x} = dx/dt-$ Newton's notation for time derivative). Anyway, the purpose of all this was to have an alternate but mysterious formulation of Newtonian Mechanics [18].

The two forms above in equation (9) are inter-related: that is, Euler's action was the integral of pdq or

$$pdx = mv\,dx = m\frac{dx}{dt}dx = m\frac{dx}{dt}dx\frac{dt}{dt} = m\left(\frac{dx}{dt}\right)^2 dt = mv^2dt = 2T\,dt. \tag{10}$$

It matches, but If we wish to "extremize" paths, a differing constant of proportionality wouldn't matter.

It wasn't perfectly clear why this approach worked or what it might really mean until it was applied to quantum mechanics and waves. The action then becomes proportional to the total number of waves along a path (or total phase), and the best path is one that provides the most constructive wave interference at the end points. So now, lets just forget some of this previous history, and work backwards to find an action and Lagrangian that allows this to happen for a single free particle.

Since $p = h/\lambda$, a wave-count along a path is $n = \Delta x/\lambda = p\Delta x/h = 2T\Delta t\,/h$, just like the transformation of pdx in the above equation (10). Nearby paths with nearly equal wave phases or counts, n, will have good constructive interference and be preferable and stationary.

To complete the counts calculation, note that total wave phase is seen in equation (2) as $\phi = (kx - \omega t) = (px - Et)/\hbar$, where E = T+V [19]. So, wave counts is:

$$n = \frac{\Delta\phi}{2\pi} = \frac{(p\Delta x - E\Delta t)/\hbar}{2\pi} = \left(\frac{2T - (T + V)}{h}\right)\Delta t = \frac{T - V}{h}\,\Delta t = \frac{L}{h}\Delta t. \tag{11}$$

[18] For more discussion on Least Action, see the Feynman Lectures on Physics [15].

[19] Remember, we are dealing with non-relativistic mechanics, so mass energy is not included. If it were, then intrinsic vibrational frequencies would be huge– almost beyond measurement.

Where the elementary classical Lagrangian is $\mathbf{L} = \mathbf{T} - \mathbf{V}$.

5. Relativity

The fundamental ideas of special relativity (SR) can now be found in any of hundreds of basic books on elementary modern physics, and those approaches will not be stressed here. Historically, the ideas for length contraction and time dilation go back at least to the works of Lorentz, FitzGerald, and Poincare on the properties of Maxwell's equations, the ether and the null result of the Michelson-Morley experiment (1887). "Lorentz Transformations" were in common use prior to Einstein, but their interpretation wasn't clear. Einstein's 1905 paper on the electrodynamics of moving bodies changed beliefs by using two central simplifying assumptions:

"The Principle of Relativity:" physical laws are invariant with respect to frames of reference in uniform motion relative to each other, and,

"The Principle of Invariant Speed of Light:" Light speed $\mathbf{V} = \mathbf{c}$ regardless of the uniform motion of either an emitter or observer.

This different point of view made the previously all-important "luminiferous aether" now seem superfluous [20]. Note that these postulates are "in the language of physics" rather than abstract mathematics. It is hoped that quantum mechanics may someday be derived from similar physical principles. Rather than the term "relativity," Einstein later wished that he had used the term "invariance" instead (e.g., the laws of physics should be invariant under Lorentz transformations – a symmetry principle). It is also implicitly understood that the space we live in is isotropic and homogeneous (and this gets carried over to cosmology as well).

Here, I would like to approach the subject of relativity and time dilation in a slightly different way beginning with the concept of metric. In geometry, we can look at distances as the positive value of the separation of two marks on a measuring tape. In relativity, we shift from "marks" in space to **"events"** which take place in both space and time (four coordinates or 4-dimensions). And we compare the separation of two events in terms of a transit of a **beam of light** between events. This is a profound difference in views. The standard mathematics for "metric spaces" insists that distance measures be positive. In relativity, we break this rule and consider both positive and negative distances and treat the signature of space differently from that of time.

[20] But Einstein changed his mind about the aether after the success of General Relativity of 1915. For him, aether was now the geometry of space-time, $g_{\mu\nu}$, later on to include other fields as well.

The important Pythagorean theorem for right triangles on a plane states $a^2 + b^2 = c^2$; for example, sides of length 3 and 4 give a hypotenuse of length 5, i.e., $3^3 + 4^2 = 5^2$. A distance between two points of 5 units will be preserved regardless of the coordinates being used. A student would be incredulous if someone claimed that the interval between these same two points instead obeyed a metric looking like $distance^2 = 4^2 - 3^2$, but something like that happens in special relativity. The usual Pythagorean idea can be extended to three-dimensions: $(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = \ell^2$ as a Euclidean (E^3) metric distance. For more general cases, we use small increment change, "$d = tiny\ \Delta$" distances, and allow arbitrary coordinate systems (e.g., cylindrical or spherical) and summarize weighted coefficients symbolically.

In conventional notation, we now write: $dx^2 + dy^2 + dz^2 = \mathbf{ds^2} = \mathbf{g_{ij}dx^i dx^j}$ with orthogonal coordinates now labeled dx^1, dx^2, dx^3 and superscripts $i\ and\ \jmath$ just standing for coordinate values 1, 2 and 3 (which could mean x, y, z). The new coefficients, g, refer to "metric tensor." So, for usual E^3 space, the metric coefficients are just trivially $g_{11} = 1$, $g_{22} = 1$ and $g_{33} = 1$, with all other indices $g_{i\neq j} = 0$. We call this a "diagonal" metric." The term $\mathbf{ds^2}$ is called an "interval" and is invariant. We can select any coordinate system in E^3 to specify the coordinates of any two given points. We can rotate and translate the orthogonal axes in any way; and the resulting distance interval Δs^2 will be the same. This concept carries over to the invariance of relativistic space-time intervals; and this can be used to simplify calculations (as shown below for length contraction).

In general relativity (GR), the idea of gravity is replaced by curvatures deriving from a general 4-dimensional space-time metric called "$g_{\mu\nu}$" shown like a 4×4 matrix of values and called a metric tensor (with "g" for gravity). We let the subscripts $\mu\ and\ \nu$ stand for values 0,1,2 or 3 where the index "0" is reserved for time and 1,2,3 for space coordinates. The g's can be functions rather than constant values and can represent curvatures of space-time. The tricky thing about both SR and GR is that the metric distance doesn't have to be positive and its components for space and time can have opposite signs! This is often hard to grasp and goes under the name "pseudo-Riemannian-metric". So, what's that all about?

The big change for special relativity is that instead of usual distances, we now care about and focus on **"light" with speed c as a fundamental reference**. If two events in space-time are connected by a beam of light (or other massless radiation), we now want their separation interval to be called **"zero!"** For the general case, the metric this time is the **difference** between time and space increments: e.g., $ds^2 = c^2dt^2 - dx^2$. A 'time-like' convention uses a plus sign on time (sign $g_{oo} = +1$) and minus sign on space, and $ds^2 = c^2d\tau^2$ where τ is called 'proper time' meaning time in the frame of a moving object. For light, $dx = cdt,\ so\ ds^2 = 0$ as desired. So, the change from classically traditional positive metrics to difference metrics is due to the change of reference to light.

Time Dilation: A particle having mass will move more slowly than the speed of light, $v < c$, and we can write our metric as

$$ds^2 = c^2\, d\tau^2 = (g_{oo} = +1)c^2\, dt^2 + (g_{11} = -1)(dx^1)^2 \text{ or:}$$

(12) $$d\tau^2 = dt^2 - \frac{dx^2}{c^2} = dt^2\left(1 - \frac{(dx/dt)^2}{c^2}\right) = dt^2\left(1 - \frac{v^2}{c^2}\right), \textit{ so } dt = \frac{d\tau}{\sqrt{1 - v^2/c^2}} \equiv \gamma d\tau.$$

This is "time dilation," and it can be picked off straight from the Lorentz SR metric form. The 'Lorentz factor' $\gamma \geq 1$; so, for example, if $v = 0.95c$, then $\gamma = 3.2$. Perceived time duration is larger than the clock time in the frame of the moving object. Then, for example, $\Delta t = \gamma \Delta\tau$ means that a muon with short half-life streaking through our atmosphere can live longer than it would at rest and be able to make it all the way through our atmosphere to the ground.

Also note that since light travels at speed v = c, the Lorentz factor is $\gamma = \infty$ so that $d\tau = dt/\infty = 0$. So, even though two events may be light-years apart, in the frame of a photon there is no advancement in time. Time flow is Zero (and, as intended, the "interval" $ds^2 = 0$). A photon leaves its source, "snaps its fingers," and instantaneously arrives at its absorber.

Length Contraction: Perhaps the simplest example of SR length contraction in the direction of motion is based on general interval invariance. Imagine a longitudinal bar in system S' of length L' moving to the right with velocity v relative to system S. Let two small flashes (events) occur when the leading and then the trailing edges of the bar coincide with a fixed post in S. $ds^2 = ds'^2 = c^2\Delta t^2 - \Delta x^2 = c^2\Delta t'^2 - \Delta x'^2$. Since $\Delta x = 0$ in S, the Δt is proper time $= \Delta\tau = L/v$. $L' = v\Delta t'$ and $\Delta x' = L'$. Then:

(13) $$(cL/v)^2 - 0 = (cL'/v)^2 - L'^2, \quad L^2 = L'^2(1 - v^2/c^2), \quad L = L'/\gamma.$$

Again, this is consistent with time dilation:

(14) $$\Delta\tau = \frac{L}{v} = \frac{L'}{\gamma v} = \frac{v\Delta t'}{v\gamma}, \quad \Delta t' = \gamma\Delta\tau.$$

The rest frame, S, sees a moving bar contracted along its length by the factor γ.

We now have the two key equations of special relativity, time dilation $\Delta t = \gamma\Delta\tau$ and length contraction, $L = L'/\gamma$. From these, the standard formulas for the "Lorentz Transformation" can be derived; and this can be used to show the invariance of the speed of light. Textbooks usually do this in reverse: use Einstein's postulates to derive Lorentz Transformations and then show time dilation and length contraction and then velocity transformations and relativistic kinematics. One can read textbooks for all of that.

The most famous formula in physics is $E = mc^2$, an idea dating back at least to Poincare in 1900 − but then only for electromagnetic fields. Einstein is generally given credit for this formula from 1905, but this and many other later publications by him either had important mistakes [21] or were incomplete [17] (he was generally sloppy about mathematics). The first complete proof of $E = mc^2 = \gamma m_o c^2$ was provided by Max von Laue in 1911. Correct kinematic derivations are now known by every college freshman in physics and will not be shown here.

Rest Mass: Energy is about the most important concept in physics. The rest mass of a particle is a fundamental vibration, $\hbar\omega_o = E_o = m_o c^2$. In special relativity (SR), we start with total (rest + kinetic) energy $E = mc^2 = \gamma m_o c^2$ and momentum $p = \gamma m_o v$, then:

(15)
$$E^2 = (\gamma m_o c^2)^2 = \frac{(m_o c^2)^2}{1 - \frac{v^2}{c^2}} = (m_o c^2)^2 \left[1 + \frac{\frac{v^2}{c^2}}{1 - \frac{v^2}{c^2}}\right] = (m_o c^2)^2 + (\gamma m_o v c)^2 = (m_o c^2)^2 + (pc)^2.$$

The same process can be repeated for frequency, $\nu = \gamma \nu_o$, and we differentiate between group velocity, $v = v_g$, and phase velocity, $v_\phi = \lambda\nu$, and the product $v_g v_\phi = c^2$. Then, we get:

$$\nu^2 = (\gamma\nu_o)^2 = {\nu_o}^2 + \nu^2 c^2 \frac{v^2}{c^4} = {\nu_o}^2 + \left(\frac{c}{\lambda}\right)^2, \; or \;\; \boldsymbol{\omega^2 = {\omega_o}^2 + (kc)^2}. \tag{16}$$

This can be conveniently pictured by right triangles having hypotenuse E with sides $(m_o c^2)$ and (pc) (or hypotenuse ν with sides ν_o and (c/λ)) [22].

Either way, if E and p are known, then m_o rest mass is also known from the wave code. And if frequency ν and wavelength λ are known, then rest frequency ν_o is also known. If $\omega/k = d\omega/dk = c$, then $m_o = 0$. So waves carry all this information even with very low amplitude. Redundantly, the knowledge of rest masses for the elementary particles is built into and accessible from the quantum fields of the Vacuum.

So, a wave enables determination of momentum or energy despite having weak amplitude, uncertainty is built into any wave-packet, and also rest frequency (or rest mass) can be deduced by the form of the dispersion relation, $\omega = \omega(k)$, which now also includes $E = E(p)$.

---

[21] The fact that Einstein's proof was not correct is detailed in the paper "Derivation of the Mass-Energy Relation" by Herbert E. Ives, Journal of the Optical Society of America v.42, p. 540 (1952).

[22] This is equivalent to the "on mass shell" 4-vector form $p_\mu p^\mu = (m_o c)^2$, or $c^2 p_\mu p^\mu = E^2 - (pc)^2 = {E_o}^2$ (also called the "mass hyperboloid" equation). Real observable particles have momentum vectors on-shell; but so-called virtual (internal Feynman line) particles have off-shell momenta.

The Schrödinger equation is non-relativistic with $KE = p^2/2m$. In that case, angular frequency would be written as $\omega = \omega(k) = (\hbar k)^2/2m + V(x)/\hbar$. Then group velocity is $v_g = v = \partial\omega/\partial k = \hbar k/m$, and phase velocity $v_\phi = \omega/k = \hbar k/2m + V/\hbar k$. Then $v_\phi = v/2 + V/\hbar k$. If the potential was not included ($V = 0$), then $v_\phi = v/2$ would seem very non-physical) and $\nu\lambda = v/2$ for the free particle. $p = h/\lambda = mv = mv_g$, so mass $m = h/v_g\lambda$. Of course, the non-relativistic case ignores the intrinsic frequency of rest mass.

Special Relativistic Lagrangian by Counting Waves: $L = -m_o c^2/\gamma - V$ appears implausibly different from the previous $L = T - V$ form. But, viewed from wave counts or total phase along a path, it becomes simple and almost obvious. We reuse the previous conversion from Euler's integral of $p\,dx \to mv^2 dt$ from equation (10). Relativistic energy is now $E = \gamma m_o c^2 + V$ (from the equations under the section on "rest mass"); and $p = \gamma m_o v$ so that $pdx \to \gamma m_o v^2 dt$. Then, wave counts becomes:

$$n = \frac{\Delta\phi}{2\pi} = \frac{(p\Delta x - E\Delta t)/\hbar}{2\pi} = \frac{-\gamma m_o c^2 \Delta t(1 - v^2/c^2) - V\Delta t}{h} = \frac{-m_o c^2 \Delta t}{\gamma\ h} - \frac{V\Delta t}{h}. \tag{17}$$

And Least Action can be written as $\delta A = \delta \int L dt = 0$, where

$$A = -m_o c^2 \int_{t_1}^{t_2} \frac{dt}{\gamma} - \int_{t_1}^{t_2} V dt. \tag{18}$$

There are many more important topics in special relativity that could be discussed here. But, now we wish to move ahead to the topic of general relativity [GRT] and see if it can be easily and intuitively approximated.

The 'Principle of General Relativity' or "Principle of Equivalence" [PE] says that a local inertial system experiencing a constant gravitational force is equivalent to a noninertial system undergoing constant acceleration (relative to the"fixed star"). The fundamental laws of physics do not depend on relative motion nor relative acceleration; they are valid for both inertial frames and noninertial frames of reference. A precursor to this is the recognition by Galileo and Newton that gravitational and inertial mass seem to be the same for all substances.

Weak Field General Relativity: Fairly simple arguments show that we can derive some first order general relativity results just using some of the arguments discussed in all of the preceding text above. We don't need the full power of Einstein Field Equations. Picture in your mind the surface of the Earth with some objects above it which we will allow to fall freely under gravity (and no atmosphere, just ideal vacuum). Consider a clock 'A' placed h meters above clock 'B' in a local gravitational field, g, with another reference comparison clock 'C' lying high but nearby at a fixed altitude [9]. The GR principle says that the physics of this system is equivalent to that where clocks A and B accelerate upwards with acceleration $a = |g|$. Then, by conservation of kinetic plus potential energy, the speeds of the clocks when they pass altitude C must obey $v_B^2 = v_A^2 + 2ah$. By SR, the clock periods dilate by $T = \gamma\tau \simeq \tau(1 + v^2/2c^2)$. Then period:

$$T_B \simeq T_A[1 + (v_B^2 - v_A^2)/2c^2] \simeq T_A(1 + gh/c^2) \simeq T_A\left[1 + \frac{GM}{c^2 r_B} - \frac{GM}{c^2 r_A}\right]. \tag{19}$$

We have used the approximation $1/r_B - 1/r_A = (r_A - r_B)/r_A r_B = h/\bar{r}^2$ and $GM/r^2 = g$. But we really can't go beyond little h distances to long radial r distances like we do in general relativity.

This period elongation, $T_B > T_A$, is called 'Red Shift.' That is, since the lower time period is longer, the perceived frequency is lower (and we say shifted toward the red). This concept has been proven to apply to both light and to 'matter waves' as well [16]. Phase difference measurements in an atom or neutron interferometer are the same as those accumulated using conventional clocks following the same paths (test of the principle of equivalence, the famous COW experiment [22]). The first accurate test of gravity on photons was the 1960 Pound-Rebka experiment for gamma-rays from Fe^{57} at an elevation of 22.5 meters above a detector (verified with accuracy $\pm 1\%$ [20]). Atomic clocks are now so accurate that a change in time flow should be seen over an elevation change of only 2 cm (Jun Ye, JILA/Nist test in Boulder, [21]). The degree of gravitational redshifting $\Delta\lambda/\lambda \sim \Delta\Phi/c^2$ is just -2 parts-per million for photons leaving the Sun but a powerful $-10\,\%$ for neutron stars [25] where the needed escape speed is 30% of light speed.

A similar comparison exists for measuring rods in the radial direction where now $L = L_o/\gamma \simeq L_o(1 - v^2/2c^2)$. Then,

$$L_B \simeq L_A[1 - (v_B^2 - v_A^2)/2c^2] \simeq L_A(1 - gh/c^2) \simeq L_A[1 - GM/c^2 r_B + GM/c^2 r_A] \tag{20}$$

If A is far away (e.g., the earth observing the sun), then

$$\frac{dt'}{d\tau} \simeq \frac{T_B}{T_A} \simeq [1 + GM/c^2 r] \quad \textit{and} \quad \frac{dr'}{dr} \simeq \frac{L_B}{L_A} \simeq [1 - GM/c^2 r]. \tag{21}$$

The term GM/c^2 is often shortened to just m. This is especially true when using modern units where basic constants are set equal to unity, $c = \hbar = \ G = 1$. The equations above can be assembled by components into a metric:

$$\mathbf{d\tau^2 \simeq dt'^2(1 - 2m/r) - dr'^2(1 + 2m/r) - {dr'}_\perp^{\,2}} \tag{22}$$

which resembles the linearized Schwarzschild metric. But this was only constructed using the principle of equivalence and special relativity for weak fields.

The arguments leading to equation (22) can be reinforced by other physical considerations. Simply by conservation of energy and basic quantum laws, a photon of energy $E = h\nu$ rising against a gravitational potential must have its frequency lowered by $\Delta\nu/\nu = gh/c^2$.

The red-shifting due to the field of our sun is a tiny contribution (e.g., parts per million). Since $\nu\lambda = c$, $d\nu/\nu = -d\lambda/\lambda$. If T is period, and ϕ is gravitational potential $-MG/r$, then [10]:

$$\frac{\nu_A - \nu_B}{\nu} = \frac{T_B - T_A}{T} = \frac{\lambda_B - \lambda_A}{\lambda} = \frac{\phi_B - \phi_A}{c^2} \tag{23}$$

Duration is a number of periods and length is a number of wavelengths. So this result is consistent with the first order length transformation (20). Massive particles also obey $E = h\nu = mc^2$ and will also suffer frequency change from change in gravitational potential.

Also notice that the radial component of the speed of light is no longer seen as constant everywhere,

$$\frac{dr'}{dt'} \simeq \frac{dr}{d\tau}\left(\frac{1 - m/r}{1 + m/r}\right) \quad \Rightarrow c' \simeq c(1 - 2m/r) \tag{24}$$

it slows down in near field. Light speed c is a local constant, but at distance separation it is non-constant and non-isotropic. This slowing down of the apparent speed of light is similar to having the gravitational field act as a refracting medium. Then light rays passing through this medium will get bent. This can be used to derive the "bending of starlight" and the "time delay of radar." So, we can get three of the simpler consequences of testing GRT. With care, the first approximation of the perihelion shift of the orbit of the planet Mercury can also be attained. This means that full testing of the Einstein Field Equations requires strong fields (such as the famous binary pulsar first seen by the giant Aricebo radio telescope [Puerto Rico, 1974]).

<u>Geodesics are world lines of extremal proper time [8]:</u> The solutions for trajectories in general relativity are curved pathways called geodesics. These are like straight lines for light rays in Euclidean space or great circles on the surface of the Earth. We wish to talk about particle paths in the gravitational field just above the surface of the Earth, and we just showed that clock frequency speeds up with height. And in special relativity, we showed that clock frequency slows down with speed. Recall that a little trick here is that frequency is 1/ clock period; they are inversely related. Now we wish to combine these. There is a time flow tradeoff between speed and elevation called "most hang-time and least speed" [Feynman] over desired trajectories. Or, in the Feynman Lectures, it is said, "An object always moves from one place to another so that a clock carried on it gives a longer time than it would on any other possible trajectory – with of course the same starting and finishing conditions" [15]. We can combine the previous math for time dilation (12) and gravity time (19) to get:

$$\left(\frac{d\tau}{dt}\right)^2 = g_{oo} - \left(\frac{dx}{cdt}\right)^2 = g_{oo} - \frac{v^2}{c^2}, \quad or \quad \frac{d\tau}{dt} = \sqrt{1 - \frac{2MG}{c^2 r} - \frac{v^2}{c^2}} = \Gamma^{-1} \tag{25}$$

where Γ could be called a "gravitational Lorentz factor" [23] and u is a velocity accompanying each potential. To low order, one could approximate this as:

$$\frac{d\tau}{dt} = \frac{1}{\Gamma} \simeq (1 - \frac{MG}{c^2 r} - \frac{v^2}{2c^2}), \quad or \quad \Delta\tau \simeq \Delta t \left[1 + \frac{g\Delta h}{c^2} - \Delta\left(\frac{v^2}{2c^2}\right)\right] \tag{26}$$

Since $\omega_t/\omega_\tau = d\tau/dt$, we do see that clock frequency (in perceiver frame) is slowed down by motion and speeded up by height. The fraction $c^2/g = R$ is the relativistic radius of curvature of the Earth's gravitational field (which works out to be about one light year)[23]. Recall in equation (16) under rest mass that frequency $\omega = \gamma E_o/\hbar = \gamma m_o c^2/\hbar$. This now becomes $\omega = \Gamma m_o c^2/\hbar = \omega_o + \Delta\omega$. This suggests multiplying the last equation (26) by $m_o c^2/\hbar$. Then, $\Delta\phi = (-L)\Delta t/\hbar$ where $L = T - V$ is just the simple Lagrangian as in the older equation (11).

Finally, returning to the principle of extremal proper time, the whole proper time accumulated along a trajectory is given by

$$\tau = \int_{t_1}^{t_2} d\tau = \int_{t_1}^{t_2} g_{\mu\mu} dx^\mu dx^\mu dt \simeq \int_{t_1}^{t_2} (1 - \frac{L}{m_o c^2})dt \tag{27}$$

In finding extremum's, added constants and proportional factors don't matter, so we are left with just the usual least action variation of the Lagrangian: $\delta\tau = 0 \Longrightarrow \delta\int Ldt = 0$. [24]

Proper time (time carried by a moving frame) τ is a maximum, and $A = \int Ldt$ is a minimum. In a way, this principle of maximum proper time is another way to derive the simplest Lagrangian, $L = T - V$ (this time for gravitational potential energy).

Cosmology: Finally, a fair understanding of cosmology can be attained simply by using Newtonian calculations for the case of a homogeneous and isotropic universe with zero curvature ($k = 0$), no cosmological constant ($\Lambda = 0$), no pressure, and only matter (like dust). From simple conservation of energy, equations can be derived resembling Einstein's general relativity field equations. This is essentially the Einstein de Sitter (EdS) model of 1932 for a "just right" universe that barely expands forever. This is discussed in many older references and was a dominant model in cosmology for nearly 50 years [24] [25]. Of course, we are missing the early radiation era of the expansion of the universe which was dominant until 47,000 years after the big bang. And we are also missing the accelerated expansion era which may have begun 7 billion years after the big bang.

This special EdS case is contained in Friedmann equations begun in 1922 which can also be approached using Newtonian conservation of energy [26]. An easy outline of essential

[23]The trajectory of a ball tossed into the air is a parabola. Change the time axis to ct, and this parabola is the approximation to the top of a really great circle of radius R (found simply by calculating radius of curvature from standard calculus formula). So, $g\Delta h/c^2 = \Delta h/R$ is a really tiny number.

[24]Feynman's derivation of this is somewhat easier than mine, see Vol II pg 42−13.

[25]Despite its historical importance, EdS is not now dominant in current books on cosmology or general relativity. Its primary utility was easy integrations of its equations for applications.

equations for Newtonian cosmology is contained in [27]. Of course, ultimately, one would wish to know the proper study of general relativity using the concept of curvature. And GRT says that the previous understanding of Newtonian gravitation is conceptually wrong and should be stated as a low order curvature of time, $d\tau/dt = g_{oo}(r)$. Beyond that, bending of light (traveling at the speed of light) sees an additional but equal contribution due to the curvature of space. This doubles the bending that Newton might have predicted.

A Newtonian approach can also be used to easily understand cosmic inflation too. Imagine an ideal case of a ball freely falling through cylindrical hole drilled all the way through the center of the Earth. Remember that the acceleration of gravity, g, only depends on the mass contained within a spherical shell at radius r. Without any air resistance, the motion is approximately that of simple harmonic oscillation with a period of 1.4 hours. Now switch from gravity to anti-gravity from the cosmological constant, Λ to give $F = +k\,r$ with $\Lambda \sim -8\pi G\rho/c^2$ in its behavior. Its solution now changes from sine-wave motion to rapid exponential expansion like $r = r_0 e^{+\sqrt{\Lambda/3}\,t} = r_o\, e^{Ht}$ [24].

References

[1] Laurent Nottale, Marie-Noelle Celerier, "Derivation of the postulates of quantum mechanics from the first principles of scale relativity," arXiv:0122.2418v1 [quant-ph] 15 Nov 2007.

[2] Leonard I. Schiff, Quantum Mechanics (Third Edition) McGraw-Hill, 1949, 1955, 1968.

[3] Steven Weinberg, Lectures on Quantum Mechanics, Cambridge, 2013, 358 pages.

[4] Atomic hydrogen in momentum space, Complement 5.F., p 442, Bucknell University QM course: http://eg.bucknell.edu/~koutslts/Ph332/Text/Chap5_JQPMaster.pdf

[5] Ron Bracewell, The Fourier Transform and its Applications, McGraw-Hill, 1965, e.g., pg. 160.

[6] Matt Visser, "The utterly prosaic connection between physics and mathematics," 2015 FQXi essay,

[7] Sunny Y. Auyang, How is Quantum Field Theory Possible?, Oxford, 1995.

[8] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler, GRAVITATION, W.H. Freeman and Co., San Francisco, 1970, 1973. [MTW]. (e.g., proper time p 315).

[9] L.I. Schiff, "On Experimental Tests of the General Theory of Relativity," Am. J. Phys. 28, 340 (1960) [free at http://www.personal.psu.edu/faculty/p/j/pjm11/AJPSchiff.pdf].

[10] David Peterson, General Relativity for Pedestrians, June 1974, limited distribution. Portions also found in my "General Physics 201-4" CU Lifelong Learning, Independent Study Syllabus 1977.

[11] H. P. Robertson, Thomas W. Noonan, Relativity and Cosmology, W. B. Saunders Company, Philadelphia, 1968.

[12] John D Norton, "General covariance and the foundations of general relativity: eight decades of dispute," Rep. Prog. Phys. 56 (1993) 791-858, Printed in the UK [also at: http://www.pitt.edu/ jdnorton/papers/decades.pdf].

[13] R. P. Feynman, A. R. Hibbs, Quantum Mechanics and Path Integrals, McGraw Hill Book Company, 1965.

[14] Richard P. Feynman, QED The Strange Theory of Light and Matter, Princeton Science Library, 1985.

[15] Richard P. Feynman, Robert B. Leighton, Matthew Sands, The Feynman Lectures on Physics, Addison-Wesley, 1964. Volume II lecture 19 is on "The Principle of Least Action." And Chapter 42 includes general relativity and maximum proper time. The whole of Volume III is on Quantum Mechanics

[16] Michael A. Hohensee, Steven Chu, Achim Peters, and Holger Müller, Equivalence Principle and Gravitational Redshift, 21 Feb 2011 [gr-qc] http://arxiv.org/pdf/1102.4362

[17] Hans C. Ohanian, Einstein's Mistakes, The human failings of genius, W.W. Norton, 2008.

[18] Robert Griffiths, Consistent Quantum Theory, 2003, free at: http://quantum.phys.cmu.edu/CQT/toc.html
[19] Free version of Quantum Field Theory Demystified at, math: http://ajbell.web.cern.ch/ajbell/ Documents/eBooks/Quantum%20Field%20Theory%20Demystified˜tq_darksiderg.pdf .
[20] Pound, R. V.; Rebka Jr. G. A. (April 1, 1960). "Apparent weight of photons." Physical Review Letters 4 (7): 337−341. And also see: Pound, R. V.; Rebka Jr. G. A. (November 1, 1959)."Gravitational Red-Shift in Nuclear Resonance." Physical Review Letters 3 (9): 439−441.
[21] J. Ye., et.al., " Systematic evaluation of an atomic clock at 2×10^{-18} total uncertainty," Nature Communications 6, Article number: 6896 doi:10.1038/ncomms7896, 21 April, 2015.
[22] R. Colella, A.W. Overhauser, S.A. Werner, "Observation of gravitationally induced quantum interference," PRL 34(23), 9 June 1975.
[23] David L. Peterson, "A Stroll Through Physics," 2014 (at www.sackett.net/DP_Stroll.pdf) under section "Decomposition of the Perihelion Shift of General Relativity."
[24] David L. Peterson, "A Stroll Through Physics," 2014 (at www.sackett.net/DP_Stroll.pdf) under sections on "The Radius of the Universe," "Cosmological Distances," and "Comments on Inflation."
[25] R. D. Blandford, "A century of general relativity: Astrophysics and cosmology," Science mag 6 March 2015 Vol. 347 Issue 6226 p 1103.
[26] Andrew Liddle, An Introduction to Modern Cosmology, 2nd Ed., Wiley, 2003.
[27] Mark Whittle, Cosmology: The History and Nature of our Universe, The Great Courses, Guidebook, Appendix Lecture 11 p 202, 2008.

6. Appendix

More Traditional Derivation of Relativistic Length Contraction:

The usual textbook calculation of length contraction is a little longer than the simple argument shown previously above [10]. It depends on a light flash mirror bounce from the end of a rigid moving bar. So imagine a longitudinal bar (say in a moving or primed S' system) moving to the right along an x-axis. A light flash is sent from the left of the bar to a mirror, M, on the right side which reflects the light back to the left end for a round-trip journey. The initial light flash event is at initial moving time $t' = t = 0$ for moving clocks versus clocks at rest in an un-primed system S. We compare length and time for the events that (0): initial flash, (1): light bounces from M, and (2) light received at the left end of the bar at final times t and t'. The length of the bar is $L' = ct'/2$; and the times recorded in the rest system are t_1 and t_2.

Since S' is moving, the time at which light hits mirror M is $t_1 = (L + vt_1)/c$. Or $t_1 = L/(c - v)$. And then the time back to O' is short because the left end of the bar has moved a total distance $x_2 = vt_2$ during round-trip transit.
Consider the last time increment $(t_2 - t_1) = (x_1 - x_2)/c$ where final $x_2 = x_1 + v(t_2 - t_1) - L$ or $(t_2 - t_1)(c + v) = L$. Then,

$$t_2 = (t_2 - t_1) + (t_1 - 0) = \frac{L}{c+v} + \frac{L}{c-v} = \frac{2Lc}{c^2 - v^2} = \frac{2L/c}{1 - \frac{v^2}{c^2}} = \frac{2L\gamma^2}{c}. \tag{28}$$

Now remember from equation (12) that the moving clock time is $\tau = t' = t/\gamma$, and $t = t_2$, and $L' = ct'/2$, so,

$$L = \frac{tc}{2}\left(1 - \frac{v^2}{c^2}\right) = \frac{tc}{2\gamma^2} = \frac{t'c}{2} \cdot \frac{1}{\gamma} = \frac{L'}{\gamma} \tag{29}$$

So, the length of a moving rod is seen as contracted by the system at rest, $L = L'/\gamma$. Notice that we have had to use the same speed of light factor c in both the S system at rest and in the moving system S'. So Einstein's second postulate of the constancy of the speed of light is still required.

Free Fall: Technically, it is not quite true that free fall in a gravitational field is the same as the effects of an observer's acceleration [12]. Real gravitational fields have tidal forces so that the Riemann tensor is non-zero. In Newtonian gravitation, tidal accelerations mean that objects at different altitudes experience different relative accelerations, $\Delta a \simeq 2MG\Delta h/R^3$. Tidal accelerations cause divergence of initially parallel geodesics in the curved space-time of GR. The equivalence principle was a guiding concept towards GR but acted as a midwife rather than actually constituting an explicit portion of GR. Nevertheless, it could be argued that PE combined with SR should produce space contraction along with red-shifting time effects and that the Schwarzschild form of a metric tensor is more physically valid than an isotropic form. The principle of general covariance would argue otherwise; but it really doesn't have a legitimate power to be convincing. For an external observer 'relatively' lacking in velocity with respect to a central mass, the radial coordinate about the central mass is 'really' different from the angular coordinates because of radial spatial contraction. And radial space contraction and time dilation only need to be approximated to first order in gravitational potential to yield the correct perihelion shift [11].

The First Test of Gravitational Red Shift: Notice that the term $-GM/r$ is just Newtonian gravitational potential, φ. In weak fields and negligible speeds, $dt/d\tau \sim 1/\sqrt{1+2\varphi/c^2} \sim 1-\varphi/c^2 = \nu_o/\nu$. If ν is light frequency (the inverse of light period), $\nu(r) \sim \nu(r_o)(1+\Delta\varphi/c^2)$. On the surface of the earth, $\nu(h) = \nu(h_o)(1 - g(h-h_o)/c^2\ )$. This important 'red shift' of light at different potentials has been verified experimentally even over short altitude changes on Earth [e.g., within ±1% for the 'Pound-Rebka' experiment over $\Delta h = 22.5\,m$ back in 1959 [8]] [26].

Some comments on going Beyond Non-Relativistic QM to Quantum Field Theory (QFT)

A main difference between QM and QFT is that at higher energies, the number of particles present is not conserved. Matter and radiation are easily inter-converted (as long as appropriate quantum numbers are preserved). Key new operators are then introduced beyond those of relativistic QM: creation operators and annihilation operators (called $\hat{A}^\dagger$ *and* $\hat{A}$) and are related to the raising and lowering operators for the energy levels of the Linear Harmonic Oscillator (LHO). This is very different from ordinary (non-relativistic) quantum mechanics where we discuss the evolution of a "particle" already in existence with particle number being held constant. Psi is not a probability amplitude but

[26] Actually, weak field red-shift can be derived without General Relativity by simply using the principle of equivalence and special relativity (see Schiff [9]).

operators which create and destroy particles in various normal modes.

Fundamental reality is composed of fields. And there are two basic types called "matter" fields and "interaction" (or "gauge") fields, and they have quanta for fermions and bosons (half-integral spins and integral spins). Fundamental interactions occur only between matter and interactions fields [7]. "So, is QFT really based more on particles or on fields? Although there is still a little disagreement, a strong majority of theoreticians favor fields as fundamental objects. Nature is made of fields. Quantum fields permeate space-time, are relatively eternal and omnipresent, and have excited state quanta that we have traditionally called 'particles.' There is a special quantum field for each type of elementary particle.

Matter in general is an excitation or wave in one or more of the fermionic matter fields. For an electron two-slit diffraction for example, the extended singly-excited electron field goes through both slits. The interaction with a detector screen is deduced to have been from a 'particle.' "Although excitations belong to the entire field, they must interact locally." Of course, there is a problem with the word "field" in QFT (or any other classical word used to describe quantum mechanics). It is usually defined as having a value (e.g., scalar or vector, etc.) assigned to every point of space-time. We picture that simplistically as an amplitude disturbance in a mattress of springs. But the field in QFT is much more "magical" than that. Many different types of disturbances can occur at the same time in a given place and be holistically coordinated with all other locations.

The central problem with a particle interpretation is that the primary attribute of a particle should be its localization in space, and particles should be countable. But there is no such thing as an observable for position in QFT, and Wigner said in 1973 that every attempt to provide a precise definition of a position coordinate stands in direct contradiction to relativity. A 'photon' is not localizable at all, not even approximately, and there is no consistent space-time wave-function for a photon as a "particle." For single photons, one can think of an electromagnetic wave packet as a function of space-time. In general, there is no accepted viewpoint on the subject of localization in QFT that is either simple or clear even for the case of free fields. Peierls said (1973) that "at relativistic energies, the electron shows the same disease. So in this region, the electron is as bad a particle as the photon." Quantum fields are intrinsically delocalized and unbounded,?

PHYSICS LIVES IN FORM HEAVEN

DAVID L. PETERSON

ABSTRACT. Many mathematicians are Platonists in the sense of believing that major concepts and theorems are discovered rather than invented. It is claimed here that the initial foundational source of those apparently spaceless and timeless mathematical ideas is the invariant Vacuum of uniformly present space-time. This is a non-classical yet "real" Form Heaven for fundamental physics and is a storehouse for all the knowledge of the physical constants, laws, and particles of physics. The intricate structure of the Vacuum is common to all intelligences in our universe and helps to constrain the reality of their various emergent knowledge. A reductionist view begins with the basic set of quantum fields living in the Vacuum leading to more complex forms emerging from these fundamentals (protons, nuclei, atoms, molecules). These entities are quantum, and their nature along with the fundamental fields might be said to live in an unusual "square-root of reality." Mathematics applies logic, intelligence and abstraction to world patterns and then generalizes at will forming abstractions of abstractions. But the field of mathematical-physics continually cross-fertilizes math and physics modestly limiting their divergence.

1. Introduction

Focus for a moment on a simple question, "Does pi (π = C/D) exist before we discover it? And if it does, where does it exist?" Historically, our knowledge of this basic constant comes from performing measurements in our various environments and doing practical calculations. We use pi when we deal with circles, circumferences, and areas or volumes of spheres or cylinders. And then, later on, after much development, pi can also appear from a multitude of other activities such as the summing of series. The classical physical world hints at pi in many ways: Nature has spherical planets and stars, planetary orbits, periodic vibrations, spherical droplets of water, and progressing phases of waves as examples. No matter where they are, intelligent creatures trying to understand the universe will find pi useful and intriguing. In itself, Nature doesn't explicitly or overtly know pi, but Nature can be codified by rational communities trying to understand the patterns of Nature.

The Greek philosopher Plato ($\sim$ 424 - 348 BCE) stressed the importance of relatively spaceless and timeless abstract ideas relating to numbers, geometry, nature and ethics and why these concepts seemed to be universal over the world then known to the Greeks. The ability of different people to independently re-discover or "instantiate" some of these apparently

Email: davepeterson137@gmail.com. Date: 11 February, 2015, Essay updated to 26 April, 2015.

Posted at **FQXi** Foundational Questions Institute Essay Contest, 20 February, 2015, http://fqxi.org/community/ forum/category/31424 .

pre-existing ideas was believed to be due to "remembering" them from a time before birth when souls had contact with ideas in Plato's heaven. We now tend to view this idea as silly mysticism except for the nagging awareness of its general popularity across more than two millennia and its widespread popularity among respected mathematicians up to the present day. Contributing to their belief may be the desire to think that mathematical progress is indeed discovery rather than invention. We wish here to update and apply Plato's idea to modern physics and mathematics and make the "problem of universals" into "the universe."

The set of all invariant abstract ideas could be called Plato's "Form Heaven." In our modern era, the important concepts or forms of mathematics and physics have complexity and invariance that goes far beyond anything Plato could have imagined. The original intent of his Forms was that they be abstract properties, that they transcend particular instantiations, that they be pure and "perfect models" or causal templates, that there is some sense in which they are objectively "real," and that they have some sort of a hierarchy of connectedness down to instantiated objects [1]. They should not have particular places or times of existence, they are beyond localization in space or in time. The basic mathematical forms of Plato's day are still important to us today: numbers, perfect circles, spheres, geometric "platonic" solids, lines, triangles and the Pythagorean theorem. But we have now gone well beyond that. The ranking or value of a Form should reflect its degree of invariance and how generalizable it is from a fundamental reductionist sense. We may no longer consider examples like people, dogs, the color green, hair, wood, or air as being invariant enough to call Forms because some of these may be restricted too narrowly to Earth and its biology and culture. We want much wider invariance than that; and, in the following, we wish to broaden the concept of the invariance of Forms from the ancient Greek world up to the presently known universe within our particle horizon. Although concepts like reincarnation and the soul were popular in Greek culture, we tend to avoid them now − and, from our latest knowledge of cosmology, we would place some new limits on words like eternal or immutable. Plato's insistence that Form Heaven is not in space-time might also be loosened because modern physics now might actually identify it with the structures of the space-time Vacuum.

A standard objection to Platonism is asking the obvious question, <u>where</u> is this world of Forms? Ignoring Plato's answer, there are several possibilities. One is that it emerges and lives in the minds and culture and literature of a world community of very smart, inquisitive, international, rational, abstract thinking people (such as the mathematicians of our planet). So, mathematical Form Heaven results from "shining the light of intelligence" onto a given habitat. We may also ask the following question: Suppose there is a set of intelligent, independent, technological, alien civilizations scattered throughout our universe. Would we expect them to eventually evolve a mathematics structure and set of theorems approximately isomorphic to our own? (focusing on the most important theorems out of millions). And would their physicists eventually come up with something isomorphic to our standard models of cosmology and particle-physics? We feel slightly braver in posing such a question now that we have actually discovered nearly 2000 exoplanets [2], think that the total number of planets in our universe is extremely large (e.g., 10^{24}), and have slightly greater comfort in the Drake equation These may be standard questions for SETI (Search for Extra-terrestrial Intelligence, for "ETI"). Our knowledge is still weak, so this may be just a thought-question ("Gedanken"). But, we would probably all find it easy to believe that the equivalent of numbers like 2, 3, $\frac{1}{2}$, π, $\sqrt{2}$ and e would exist in all ETIs – they are just too useful and important to bypass. (And, for physics, we can imagine that knowledge of universal constant values for c, h, G, q_e, k_B, m_e, m_p, N_o, a_o and α_{EM} should eventually appear.) If we can agree this far, how difficult is it to further imagine that ideas like the Euler equation $e^{i\pi} = -1$ or the Pythagorean theorem $a^2 + b^2 = c^2$ will also be universal? And from the Pythagorean theorem, we can generalize metrics to $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$.

2. Mathematics

A common definition of mathematics is "an abstract representational system used in the study of numbers, shapes, structure, change and the relationships between these concepts [3]." It is an interdisciplinary language that has reason built into it; and this purity of reasoning is why Plato valued mathematics so highly. A favorite definition is from Paul Halmos, "Mathematics is the logical dovetailing of a carefully selected sparse set of assumptions, with their surprising conclusions, via a conceptually elegant proof. Simplicity, intricacy, and above all, logical analysis are the hallmark of mathematics [4]."
It is said that, "Mathematical platonism enjoys widespread support and is frequently considered the default metaphysical position with respect to mathematics" [5]. Consider the emphasis on the phrase <u>"there exists"</u> (e.g., an infinite number of prime numbers, $\exists\ \mathbb{P} \subset \mathbb{N}\,,\ \#\{\mathbb{P}\} = \aleph_o\,,\ \exists\, p \in \mathbb{P})$. We take the "existence" of these objects seriously. Can we also imagine that basic <u>proofs</u> for an infinite set of prime numbers might pre-exist in the book of the universe [6]? (e.g., Euclid's (~ 300 BCE) finite set of the first r primes followed by a new number $n = p_1 p_2 ... p_r + 1$ which may have an additional new prime divisor).

In his book, The Road to Reality, about mathematics and the laws of the universe, Roger Penrose says, "Platonic existence, as I see it, refers to the existence of an objective external standard that is not dependent upon our individual opinions nor upon our particular culture [7]." He devotes a whole section on whether the "Platonic world of mathematical forms" is real and decides that it is in the sense of the "objectivity of mathematical truth." Kurt Gödel also believed in the objectivity of mathematics, that an abstract realm existed, and that the only valid philosophy of mathematics was Platonism. He was a theist who believed that intellectual mathematical intuition is a kind of sense that enables us to perceive Platonic concepts which are really "out there" [8]. And Paul Erdős believed in the pre-existence of a transfinite Book that contains the most elegant and perfect proofs of all mathematical theorems [6].

Pure mathematicians would say that mathematics is pure math, although this emphasis only dates back to about 1800. Mathematical knowledge is only concerned with the realm of thought. It attempts to not consider direct application; but, mysteriously, the purity of one era sometimes becomes the application of a following era (e.g., number theory was once pure but now also applies to computer encryption). From the time of Karl Weierstrass, we focus on mathematical analysis and rigorous proof from axioms. Jean Dieudonné (and the highly abstract and rigorous French Bourbaki school) stated in 1962 that mathematical progress has almost nothing to do with physical applications [9]. But, twenty years later, modern physics again entered the picture. Michael Atiyah said that since about 1980, "some of the most exciting developments in mathematics have arisen from the interface with physics and particularly quantum field theory" ("QFT") [19]. The Fields Medal is sometimes called the Nobel Prize of mathematics. So far, Edward Witten is the only practicing physicist who has won this award (in 1990 for his 1981 proof of the positive energy theorem of general relativity). But eight other winners did work partly related to physics (delta functions, quantum groups, PDE's, Ising model, Boltzmann equation, renormalization, Brownian motion, and general relativity). Peter Woit says "Mathematics is a science, but it is not an empirical science. It insists on precise thought, rigor, clarity, high standards of proof and debate among an international community. New mathematics is motivated by numbers and geometry and also by theoretical physics" [10] (e.g., quantum field theory and string theory).

3. Physics

"The goal of physics is to study entities of the natural world, existing independently from any particular observer's perception, and obeying universal and intelligible rules [11]." We are aware that our physical laws, particles, and the constants of Nature are universal and invariant over space and time. So, it has always been clear that physical Forms are discovered rather than created by people. We know this largely due to the spectra of electromagnetic radiation detected from very distant sources and from the success of the standard model culminating in the discovery of the Higgs particle. And our ΛCDM concordance model of cosmology is now pretty solid largely due to the study of cosmic microwave black body radiation (CMB, e.g., via the Planck mission) [12].

Formerly, physical entities were said to 'exist' because they had mass. After Einstein, we might say they exist because they have energy equivalence (e.g., $E = mc^2$ and $E = h\nu$). We believe that photons and electric fields exist because they can deliver energy even though their mass is zero. A present concern is whether information also has any real existence. A problem in physics is that we presently seem to have two worlds: the classical world (largely composed of particles created long ago) and the "quantum world" (currently either coming into being or the not yet energized forms of the Vacuum). Plato's Forms were originally conceived as beginning in abstractions from the classical world where we now speak of Newtonian mechanics and gravitation applied largely to macro-bodies. The quantum world should perhaps really be called a "pre-quantum" world because its equations stop short of the actual transfer of quanta. It is a strange non-classical existence possibly described as a sort of "square-root of reality" discussed more below. We refer to this world as "real," but that is a horribly overused word that should better be called quantum-real or "qureal" instead to separate it from a usually understood bias of being classical.

Apart from solid-state physics (condensed matter), physicists generally have a reductionist perspective: work towards the bottom and then build up from there. Elementary particles and fields represent the present bottom rung of this ladder. Our past history biases us towards visualizing classical particles when thinking about 'particle physics.' But fundamental particle physics is now discussed in books on 'quantum field theory" (QFT). The belief in "a pure fields view" has developed during the past three decades. "At the high energy end, most quantum field theorists agree for good reasons that relativistic quantum physics is about fields and that electrons, photons, and so forth are epiphenomena, namely excitations (waves) in the fundamental universal fields" [15]. Quantum fields exist in space-time; but we need to talk about the nature of that existence. The reason that all electrons are the same ($N_e \sim 10^{80}$) is that they are all excitations of the same pervasive electron field. A general view is that a quantum field is an entity existing at each point of space which regulates the creation and annihilation of particles − one field for each type of particle. QFT treats fields as the knowledge embedded in the Vacuum of how to make any particle providing that adequate energy and quantum numbers are available to do so [16]. Some say that even in usual QM there is really 'no evidence for particles' [14] [15].

Frank Wilczek noted that a new term was needed which is broader and more relevant to physics and QFT than the old ideas of aether, plenum, substance, vacuum, space-time, or world-stuff [17]. He uses the word **'Grid'** as a "multilayered, multicolored cosmic superconductor" including quantum fluctuations, a superconducting condensate, a weak superconducting Higgs condensate, Einstein's metric field ($g_{\mu\nu}$), the dark energy cosmological constant grid density (Λ), and "chiral symmetry-breaking condensate consisting of quark-antiquark pairs." It is recognized that general relativity is really an "ethereal theory of gravitation." Grid superconductivity gives masses to particles created by weak bosons, and particles are relatively localized disturbances in the Grid. Some might add that the smoothly distributed cosmic black-body background (CMB) is also a modern version of an aether with a locally preferred frame corresponding to the expanding cosmic fluid. Wilczek's picture is further encouraged by the experimental finding of the 125 GeV resonance appearing to be the standard model Higgs boson (CERN-LHC, 4-July, 2012).

As a recent example of the hidden "causal templates" of the space-time Vacuum, consider particle-antiparticle colliders producing what might be called "pure energy" which in turn can then lead to myriad possible output particles of precisely defined types apparently emerging out of the Vacuum itself. Since the earth rotates and orbits, the real historical set of collision points of these colliders have been sweeping out corkscrew paths covering large samples of space and over a long time implying that this production is spaceless and timeless. There is a beautiful plot released by CERN LHC showing quark-mesons produced by the Vacuum as seen by an increasing total mass of di-muons, $\mu^+\mu^-$, "A Lovely Dimuon Mass Spectrum" [18] . The spectra of events per GeV begins with lower energy at left showing spikes in cross-section for production of mesons called η, ρ, ω for $u\bar{u}, d\bar{d}$. Then there are the unflavored quarkonia $q\bar{q}$ mesons: the ϕ meson for strangeness $s\bar{s}$, and then charmonium J/ψ for $c\bar{c}$ followed by Υ or $b\bar{b}$ and its excited states. Finally there is a huge spike for the neutral weak Z^o boson near 92 GeV. These particles are spewed forth when the Vacuum is stimulated. From a separate reference, a very similar plot of particle production cross-section also results from electron-positron e^+e^- collisions [13] additionally showing a high energy hump for W^+W^- production. We would believe that pumping energy into any point in the universe would enable the production of these same particles and deduce that the Vacuum of space-time holds the pre-existing knowledge of all these particles and more.

Plato would also not have anticipated the world of identical particles. All of the universe's muons are the same, each of its protons is the same, each of its ground state gold atoms is the same (isotope with neutron number say at 118). When we instantiate a physical form, it is not an impure poor-copy; it is as pure as the abstract forms themselves. And, experiments also show that these objects are quantum too − at least up to macro-molecule size like C_{60} carbon buckyballs. They are 'de-localized' entities. An experiment in 2013 [25] demonstrated nano-particle de Broglie matter-wave interference of macromolecules above 10,000 amu! It is not presently clear where a dividing line may be between these quantum objects and so-called classical objects. Could it be that the physical world is all-quantum? Or might an upper mass limit be near the "Planck mass" ($\sim \mu$ gram)?

4. Quantum Mechanics and "The Square Root of Reality":

Quantum mechanics provided us a strange new world where "reality" became hard to define and complex numbers became a necessity. After eighty years, there is still an intense on-going debate about the nature of the quantum state, ψ (is it "ontological" or "epistemological" or perhaps some blend of both?). As an interesting example, suppose that a minority view called the "Transactional Interpretation" has some validity [21]. In this TI world, ψ is an "offer wave" from an emitter to possible absorbers. A confirming wave ψ^* goes back in time from an absorber to the emitter resulting in a handshaking "transaction" with weight $\psi^*\psi$ which provides an explanation of the Born Rule. In this picture, the reality of a quantum state or wave function ψ is something like the "sound of one hand clapping." That is a very unusual kind of "reality," and the Form Heaven of Physics has a reality similar to this.

Take the Born rule seriously as having sub-quantum-real ('qureal') wavefunctions needing to be 'squared' to become classical candidate entities. Classically recognizable probability may be given by $P = \psi^*\psi$, where psi lives in a new sub-world resembling the pulling apart of classical reality into two "square-root" (or 'star-root') complex number parts. So, electron spin as classically real or vector-like fails to agree with observation, but quaternions or gamma matrices fit needs better. Discussions of the Born rule go from wavefunction to detection probability with a selection criterion that is unspecified and likely random. Here we wish conceptually and heuristically to go backwards, from classical to sub-quantum.

As a first example, one occasionally used representation of a single photon (the 'Riemann-Silberstein' form) is found by taking the "square-root" (or "star-root") of its supposed energy density: $\psi^*\psi \propto (\epsilon_o/2)(E^2 + c^2B^2)$ becomes $\psi = \sqrt{\epsilon_o/2}\,(E \pm icB)$ [27]. The 'star root' operation is of course not unique and not well-defined, it is intended to only be heuristic: star-root ${}^*\sqrt{Prob} = P^{*/2} = \psi$. In a similar vein, the 'Dirac program' essentially derives from taking the square root of the d'Alembertian [20]; or we can consider the Dirac equation as the 'star root' of the Klein-Gordon equation: $Dirac = (KG)^{*/2}$, i.e.,

$$(1) \qquad \left[\partial^\mu \partial_\mu + \left(\frac{mc}{\hbar}\right)^2\right]\psi = 0 \qquad \xrightarrow{*/2} \qquad i\hbar\gamma^\mu\partial_\mu\psi - mc\psi = 0$$

where the γ^μ 's are the 4×4 unitary Dirac matrices. Instead of just complex conjugation (or *'starring'*), higher dimension quantum spaces can use matrices with conjugate transpose or 'Hermitian Adjoint,' $A^\dagger$. The 'star-root' idea can go further, for example into the realm of quantum cosmology with supersymmetry where it is said that supergravity($N = 1$ SUGRA) naturally provides a Dirac-like 'square-root' of gravity [23]. And some programs for unifying general relativity and quantum mechanics use "tetrads" which can be thought of as the square-root of the metric, $g_{\mu\nu}$.

Hypercomplex numbers can have convenient application in the classical world with examples including quaternions ($\mathbb{H},\ basis = \{1,\ i, j, k\}$) for 3D-rotation, electromagnetism, and relativity. And their use eventually led to the development of more conventional vector analysis which was easier to use. But the use of hypercomplex numbers becomes a necessity in the quantum world. The algebra of hypercomplex quaternions and Dirac matrices are examples of Clifford algebras (e.g., $C_o = \mathbb{R},\ C_1 = \mathbb{C},\ C_2 = \mathbb{H},\ C_3 = Pauli,\ C_4 = \gamma' s$) .

Similarly, the physics of fields usually begins with a Lagrangian written in terms of energies and interactions. These in turn contain "squares" of fields such as the free gauge field of electromagnetism and the gauge part of the weak action [24]:

$$(2) \qquad \mathcal{L}_{EM} = -\frac{F_{\mu\nu}F^{\mu\nu}}{4} = \frac{(E^2 - B^2)}{2}, \qquad \mathcal{L}_{weak-W} = -\frac{W^a_{\mu\nu}W^{a\mu\nu}}{4}$$

Again, what is E_i or W^a_ν all by itself? Well, the fermion-boson interactions mean something: $\mathcal{L}_{int} = -J^\mu A_\mu$ [e.g., like the eA part of the Aharonov-Bohm (momentum phase change) effect; but unless it is electromagnetic frame dragging, it is hard to interpret in words].

5. Conclusion

We can now identify Plato's Form Heaven for fundamental physics with something more tangible than previous 'spirit or mental worlds.' However, these Forms of physics are objectively real in space-time in a strange way. They are the information and potential to create elementary particles; but, without energy, they are not the particles themselves. The knowledge of this world can be plucked out by hitting the Vacuum with pulses of energy. Unlike Plato, the resulting instantiations of each type of Form are themselves identical and pure as are the contents of the Vacuum itself. But, until detected, these instantiations are quantum objects lacking classical reality. The objects and the Forms have a new type of existence similar to the "square root of reality." Quantumness is preserved up to the size of macro-molecules. Beyond that, it is undecided if larger objects are classical or not ("macrorealism").

Many mathematicians generally believe that their basic theorems and concepts are pre-existing Platonic Forms which are discovered rather than invented. To some degree, mathematics is abstracted from physical reality because the regularity, repeatedness and symmetry of Nature is fruitfully expressible in the language of mathematics. There are regions of overlap between math and physics, and this overlap region of mathematical-physics has to be compatible with both. Lack of compatibility can lead to a modification of one side or the other. Historically, physics and pure mathematics are relatively free to diverge and grow apart. But then, unexpectedly, the evolving physics finds that some previously pure math can be usefully applied. And the math finds that new physics has some aspects that deserve to be better explored mathematically (and such development can be better funded). They cross-fertilize each-other. They find that they are not entirely separated but can play together.

What motivates mathematicians and physicists to devote their lives essentially to the study of these Forms? Transcendence and connectedness. We sense that we are participating in a huge world beyond our own limited experiences. We sense that the intelligences in the universe might discover the same truths we value; so we have a cosmic sharing. Without overtly expressing it, the physicist senses Einstein's "Cosmic Religious Feeling" [22] which can be essentially summarized as rational "Deep Nature Appreciation."

References

[1] David Banach, "Plato's Theory of Forms," 2006, http://www.anselm.edu/homepage/dbanach/platform.htm
[2] René Heller, "Better than Earth," Scientific American, Jan 2015, p. 34.
[3] http://en.wiktionary.org/wiki/mathematics.
[4] P R Halmos, Mathematics as a Creative Art, The American Scientist, 56, 375-389, 1968. Or see: http://www-history.mcs.st-and.ac.uk/Extras/Creative_art.html .
[5] Julian C. Cole, Platonism, Mathematical, "Internet Encyclopedia of Philosophy." http://www.iep.utm.edu/mathplat/

[6] Martin Aigner, Günter M. Ziegler, Proofs from THE BOOK, 2014 (The book is dedicated to the mathematician Paul Erdős, who often referred to"The Book" in which "God" keeps the most elegant proof of each mathematical theorem).
[7] Roger Penrose, The Road to Reality, A complete guide to the laws of the universe, Alfred A Knopf, N.Y., 2005. (Section 1.3, p 12).
[8] Hilary Putnam, Reason, Truth and History, Cambridge, 1981 (p 145).
[9] Michael Monastyrsy, Modern Mathematics in the Light of the Fields Medals," A K Peters, Ltd., 1998.
[10] Peter Woit, Not Even Wrong, the failure of string theory and the search for unity in physical law, Basic Books, 2006. P 258.
[11] Alexia Auffeves and Phillippe Grangier, "Contexts, Systems and Modalities: a new ontology for quantum mechanics," arXiv:1409.2120 [quant-ph] 23 Jan 2015
[12] Planck 2015 results. XIII. Cosmological parameters , Planck Collaboration, arXiv:1502.01589 [pdf] Also see: Planck Science Home Team Publications, http://www.cosmos.esa.int/web/planck
[13] Donald H. Perkins, Introduction to High Energy Physics, 4th Edition, Cambridge, 2000, p. 145, e^+e^- annihilation to hadrons.
[14] Casey Blood, "No Evidence for Particles," arXiv: 0807.3930, also see 0912.2985.
[15] Art Hobson, "**There are no particles, there are only fields**," arXiv 1204.4616, 52 pages, or physics.uark.edu/hobson/pubs , Submitted to AJP, 18 April 2012.
[16] Mark Whittle, Cosmology: The History and Nature of Our Universe, The Great Courses guidebook, 2008, e.g., pg. 168.
[17] Frank Wilczek, The Lightness of Being, Mass, Ether, and the Unification of Forces, Basic Books, 2008.
[18] "A Lovely Dimuon Mass Spectrum," CMS and ATLAS results from CERN, http://www.science20.com /quantum_diaries_survivor/blog/atlas_vs_cms_dimuon_resonances-75594 .
[19] Michael Atiyah, Book Review, "Bourbaki, A Secret Society of Mathematicians," Notices of the AMS, Volume 54, Number 9, p.1150, http://www.ams.org/notices/200709/tx070901150p.pdf
[20] Theodore Frankel, The Geometry of Physics, An Introduction, Second Ed., Cambridge, 1997, 2004, p. 502, Section 19.2d. "The Dirac Program: The Square Root of the d'Alembertian."
[21] John G. Cramer, "The Transactional Interpretation of Quantum Mechanics," Reviews of Modern Physics, Vol. 58, No. 3 July 1986, pp. 647-687.
[22] Albert Einstein, The World as I See It, 1949, later called Ideas and Opinions (1954), NY, Bonanza Books.
[23] Paulo Vargas Moniz, Quantum Cosmology - The Supersymmetric Perspective - Vol. 2: Advanced Topics, Springer, 2010.
[24] Michio Kaku, Quantum Field Theory, Oxford, 1993.
[25] Sandra Eibenberger, Stefan Gerlich, Markus Arndt, Marcel Mayor, Jens Tüxenb, "Matter−wave interference of particles selected from a molecular library with masses exceeding 10,000 amu," Physical Chemistry Chemical Physics Issue 35, 2013.
[26] Sundance Bilson−Thompson, Deepak Vaid, "LQG for the Bewildered," arXiv:1402.3586 [gr-qc] 1 Feb 2015.
[27] Iwo Bialynicki-Birula and Zofia Bialynicka-Birula, "The role of the Riemann-Silberstein vector in classical and quantum theories of electromagnetism," 28 Dec 2012, http://arxiv.org/pdf/1211.2655.pdf

WHAT FUNDAMENTAL SHOULD MEAN

DAVID L. PETERSON

Abstract. The quest for future deep foundations of physics should continue to pursue greater unifications, should incorporate "hypercomplex numbers" or the name "Clifford algebras" in describing its quantum realm; develop a language for discussing in what way "quantum-waves" could be considered "real;" and should always be capable of expressing the formulation and interpretation of any fundamental theory so that humans might believe it is isomorphic to Nature's actual mechanisms. That is not yet the case for present-day quantum mechanics nor quantum field theory. In addition, there is probably a limit to testably-assured knowledge perhaps three to six orders of magnitude in particle energy above present capabilities. If "ultimate reality" lies beyond that, we will never have confidence in identifying it.

1. Introduction

The entrance to my office has always had a cute picture of a family of curious Kalahari meerkats peering out upon the world. And below that is an Einstein quotation that says, "We shall never cease to stand like curious children before the great mystery into which we were born [1]." Many of us love physics and desire to know its deepest mechanisms that "bring us closer to the Secrets of the Old One."

Science has been solving the mystery of what the world is and how it works for hundreds of years. Physics had made astonishing progress in its approach towards "ultimate reality" resulting in a unification of strong, weak, and electromagnetic interactions called the "standard model" or "SM." But, this model is not yet at the "bottom of it all." There are still many outstanding problems and, by one estimate, 26 free parameters [2]. Ideally, a model with deeper foundations would uniquely specify the values of those constants. A proposed model below SM uses"supersymmetry" with over a hundred free parameters — so this also can not be "the foundation." Then there is"String Theory" or "M-Theory" that appears to be forever beyond testing and hence are not "theories" nor what we have been calling "science." A scientific theory must have a history of strong experimental support to provide a "high level of confidence." We should consider the possibility that a barrier or limit to assured human knowledge may exist.

So, one should step back from these frontiers onto an overlook and assess where we are— what we know— and how this should affect future foundations:

In the following, we focus mainly on the different levels of "unifications" of theories of high-energy physics. These increasingly deeper quantum foundations use a progression of hypercomplex numbers of increasingly higher dimension, n, that may be labeled by the name "Clifford algebras"

Email: davepeterson137@gmail.com. Date: 13 December 2017 updated to 14 January, 2018. FQXi.org Essays on "What is Fundamental." Posted as: https://fqxi.org/community/forum/topic/3027 .

$C\ell(n)$. There is no intention here to actually use these algebras beyond just providing a convenient name for the degree of hypercomplex numbers representing the progression from the real to the complex to the Pauli matrices or quaternions to the Dirac matrices to the next deeper levels. All of these levels of algebras use n base elements whose squares are either $+1$ or -1, so one might say that the quantum world below classical reality lives in "the square root of the reals:"

The progression for $C\ell(n)$ can go like this:
$C\ell(0)$ = the Real numbers, R; $C\ell(1)$ = the Complex numbers C; $C\ell(2)$ = the quaternions H (for Hamilton); $C\ell(3) = Pauli\ matrices$; $C\ell(4) = Dirac\ gammas$; $C\ell(6)$ = the Standard Model, SM; and $C\ell(7)$ = the "octonians" which use seven complex numbers.

2. Foundations of Physics: View from Above

2.1. **Reductionism has been incredibly successful for particle physics.** In some ways, it is like those nested Matrushka dolls that maybe end with a smallest last doll. The bottom doll might have information enabling calculation of the properties of larger dolls (for example the progression down in size from molecules to atoms to nucleons to quarks). Unifications are another type of reduction — pulling together two previous theories into one new theory (see examples in the section on Unifications). From the present Standard Model, we are supposed to be able to calculate masses of hadrons from colored quarks and gluons — and we can now approximate hadrons using supercomputers for "lattice-QCD" [quantum chromo-dynamics on a 4D lattice [4]]. We are supposed to be able to predict water from quantum mechanics —and again, it can be approximated using supercomputers [5].

In practice, of course, it is often wiser and easier to use emergent basics for complex systems above particle physics — use the ideal gas law for weather. And the metallic state is an emergent phenomenon. Condensed matter physics encourages emergence with the slogan "More is Different [6]". But here, we will focus mainly on "fundamental" high-energy physics.

2.2. **Particle physics is quantum.** A problem with "Ultimate Reality" is the word "Ultimate" and that pesky word, **"Real."** The Born rule says that we go from the wave-function, ψ, in "quantumland" [3] to statistical events in our measured classical reality by "squaring psi" , $|\psi|^2 = \psi^*\psi$. That suggests that psi itself lives in a different sub-world below classical "reality" with complex numbers for this deeper reality: "Quantum characteristics are irreducibly complex, they cannot be decomposed into real and imaginary parts [14]." This realm of the sub-real uses layers of fundamentality with each level inwards or down being more fundamental than the last. Progressing downwards from usual complex wave-mechanics, we can include electron fermion spin by making use of quaternions (the first hypercomplex number system [1843]) or Pauli-matrices [1927]. In 1928, the next step down resembled taking the square-root of the d'Alembertian [17] or "square root of the Klein-Gordon equation" [see Eqn. 1] resulting in Dirac spinors and 4×4 "γ"-matrices which themselves can be composed of Pauli matrices. Dirac theory is the foundation of quantum electrodynamics (QED), and such quantum field theories are intrinsically hypercomplex making them difficult to explain in words.

Summarizing this history, some key developments in modern quantum physics have been aided by taking complex and deeper "hypercomplex" numbers seriously— perhaps even as representing something somehow isomorphic to what Nature actually does. Presumably, all relevant quantum fields and $C\ell(n)$ bases are overlaid together into a collective "GRID" [22] that is in or of the "Vacuum."

Examples of the "square-root" of real numbers include: the complex numbers of quantum mechanics using the imaginary number i with $i^2 = -1$; non-relativistic electron spin uses 2×2 Pauli matrices σ_i with ${\sigma_i}^2 = +1$ related to quaternions q with bases $\{e_i\} = \{1, i, j, k\}$ with $e_{i>0} = i\sigma_i$ and ${e_{i>0}}^2 = -1$ times the identity matrix. Dirac 4×4 gamma matrices for quantum-electrodynamics can also be written as 2×2 matrices themselves containing Pauli matrices, and ${\gamma_i}^2 = +1$ unit matrix.
To avoid traditional confusion, "real" in physics is most often used to mean classical, and quantum mechanics (QM) processes the "sub-real" (quantum-real, qu-real, or Ruth Kastner's term "potentia" for potentially real prior to measurement [3]). If there isn't a name for the underlying quantum realm, few will pursue searching for the base mechanisms of it.

2.3. **Quantum Waves should be considered to be "real".** — or should I now say "sub-real?" In quantum philosophy, an observed system is real if its properties are intrinsic and observer independent. Many would say that kind of "real" doesn't quite fit. We don't know what actually goes on in quantumland, but the "observer" or "absorber" is a key part of it. Quantum mechanics began with discrete "action" [Bohr atom, 1913] along with $E = h\nu$ and with $p = h/\lambda$. Waves transmit information between emitters, interactions, and absorbers: energy is represented by the density of waves in time, and momentum is represented by the density of waves in space. Planck's constant of action h is a conversion constant that seems tiny because our systems of units are designed for big people. But, the question is always asked, "Waves of What?" It is a good question with a typical answer, "waves of probability amplitude" — ? We should be digging deeper into the sub-world.

Which Waves? Waves have a phase velocity, $v_{phase} = v_\varphi = \lambda\nu$. For free non-relativistic Schrödinger waves, $E = h\nu = KE$, and $v_\varphi = \lambda\nu = v_g/2$ where $v_g = v_{group} = v$ is the speed of the particle and its wave-envelope. These phase waves are often shown in freshman textbooks, but they are wrong because they do not satisfy special relativity. Particle rest mass should be included in the total energy, $E = h\nu$ where $\nu^2 = {\nu_o}^2 + c^2/\lambda^2$, and $\nu_o = m_o c^2/h$. For an electron, $\nu_o \sim 10^{20}$ Hz (hundred Exa-hertz, unfortunately beyond present measurement ability). Since particle speed $v_g < c$, $v_\phi = c^2/v_g > c$! – superluminal, but considered ok because phase velocity itself carries no energy. Perhaps treat it as just information being transfered in the sub-real in some vicinity near a moving particle.

How did complex variables enter non-relativistic Schrödinger wave mechanics? Begin with the "reasonable quantization axioms" $E = h\nu = \hbar\omega$ and $p = h/\lambda = \hbar k$ so that a wave has phase $\varphi = kx - \omega t = (1/\hbar)(px - Et)$. Complex numbers then entered for convenience via Euler's formula, $e^{i\varphi} = \cos\varphi + i\sin\varphi$. Take the real p and E out of the exponent by using (complex) derivative operators: $\hat{p} = -i\hbar\partial(wave)/\partial x$ and $\hat{E} = +i\hbar\partial(wave)/\partial t$ and apply these operators in $\hat{E} = \hat{p}^2/2m + V$ operating on wave $\psi(x,t)$ [19]. Complex numbers are now in the "formulation" for simplicity but later became a necessity when describing electron up-down spin with Pauli matrices (which themselves are already hypercomplex). The Schrödinger equation is a low energy case of the relativistic Klein-Gordon equation where E now includes mass energy.

The principle of least action seemed slightly mysterious in Newtonian mechanics. But in quantum mechanics action, $S = \int Ldt$, seems to be just proportional to a phase or wave counter expressed using a "Lagrangian, L" along a path with parameter t for time. Least action can give greatest constructive wave-interference and can itself be deduced from Feynman's "Sum over Histories" or "Path Integral" program.

Action for tiny particle masses may use the Lagrangian $L = KE - V$ which can count waves along a classical trajectory (like free fall from a cliff). For relativity, the Lagrangian $L = -m_o c^2/\gamma - V$ is also shown to be wave counting [19]. And, this also works for electromagnetism (EM) having a term $(\vec{p} - e\vec{A})$ where $\vec{A}$ is EM vector potential and $e\vec{A}$ is like an "EM momentum." The Standard Model Lagrangian is a complex beast having the symmetry of the standard model group: "$SU(3)_C \times SU(2)_L \times U(1)_Y$." So far, all theories make use of a Lagrangian suggesting the importance of waves for each foundation level. Instead of the phrases "matter waves" or just complex waves, there are also hypercomplex waves (!):

Schrödinger waves are complex but limited — non-relativistic and not for spin. Dirac waves are for relativistic spin $^1/_2$ fermions like the electron. "Proca" ("heavy-photon") waves include mass but are still for the electromagnetic vector potential, $\vec{A}(x,t)$, U(1) waves. Gravitational waves are waves OF space-time. There are SU(2) non-linear"Yang-Mills" wave equations. QCD with SU(3) color symmetry has waves. The bottom line is that quantum waves are waves in each relevant quantum field at different hypercomplex levels. For the case of quaternionic waves, there is a geometric phase angle in a sort of Euler polar form expressing a periodic quaternion with a real part and three imaginary parts [23]. Newly popular Geometric-algebra with $C\ell(4)$ or $C\ell(1,3)$ will cover a lot of these case.

People have been reluctant to accept the "reality" of waves for several reasons: one is just a philosophical view that nothing about the sub-real has any "existence," another is gauge flexibility so that a variety of mathematical forms will still yield the same results. And the experimental world can only detect phase "differences" and never any absolute phase. But maybe Nature actually makes particular choice selections despite our inability to see it.

The importance of quantum waves, Lagrangian wave counting, action, and Lie symmetry groups should be expected to continue through the next few generations.

2.4. **Foundations should satisfy Humans.** — Fundamental formulations should facilitate calculations but should also alternatively be expressible in a form that humans can believe might represent "How Nature actually does it!" And Humans should be able to tell each other stories about it — after all, we are intrinsically culture accumulating and "Story Telling Apes" [9].

As example, consider the case of the Theory of General Relativity: Special and General Relativity are based on a few reasonable principles (invariant laws, light speed invariance, the principle of equivalence) that are then developed mathematically. This field has grown and improved for over a century well enough that a graduate student shouldn't have much trouble, and a serious layman at least can learn "weak field" gravitation with some ease. With our new knowledge that the universe is spatially "flat," the Einstein field equations are easily derived from Newtonian arguments and understood with just a little algebra and calculus. Experiencing this realm through an introduction to modern cosmology is very fulfilling for all concerned parties [20].

Quantum Field Theory is another story [14]. The clarity, math, interpretations and even definitions [24] are difficult, and the serious layman can only see its surface from popular writings (of which there are many to see). Some books for the layman even discuss gauge symmetry, Lie groups and spontaneous symmetry breaking [25]. But with the formal math and abstract postulates, even basic quantum mechanics (QM) is challenging, lacking in visual pictures and probably will make only modest intuitive sense for most students. Wanting to believe that this is perhaps how Nature actually works just doesn't go together with saying that "A quantum state is completely specified by a vector in Hilbert space." Some may appreciate the logic of the Copenhagen postulates formalism and its mathematics, but it will not yet satisfy most humans; and no one yet knows how to make that happen. In this sense, present QM postulates and philosophy are a bottleneck against comprehending any deep reality about how Nature works. And deep and deeper fundamental physics should not be understood only by select best genius professionals nor perhaps just by some future supercomputing artificial intelligence (— will we really be able to choose not to go that far . . . ?).

3. UNIFICATIONS:

We are concerned here with present and future physical theories having different levels of fundamentality. Newtonian mechanics and dynamics used to be fundamental. New sciences were formulated with foundational principles for each: optics, electricity, magnetism, gravitation, thermodynamics and more. Maxwell's electromagnetism (EM) unified electricity, magnetism and light and thus became more fundamental. Electro-weak theory unified the weak and EM interactions. Relativity brought space and time together with that immortal quote: "Henceforth space by itself and time by itself, fade away completely into shadow, and only a kind of union of the two will preserve independent permanency" [Hermann Minkowski, 1908 [7]]. In 1960, fundamental or "elementary" particles mainly meant protons, electrons, and neutrons [1932], along with a few pions [1947] and the neutrino [detected in 1956]. Now there are many elementary particles. Protons and neutrons are now composite from colored quarks and gluons — and deriving neutrons and protons and heavy hadrons from fundamental theory is achievable but is also really hard [4].

Historical unifications included: Identifying Earth and Sun as a planet orbiting a star. Rest and uniform motion (Galilean relativity). And Sun gravity and Earth gravity beyond falling apples (universal gravitation)
Unification of theories is a measure of the degree of fundamentality.

- Electricity and Magnetism (E + M = EM) and then Maxwell's EM and Light [1862-1873]
- Space and Time (special relativity, SR, [1905])
- Acceleration and gravity (Principle of Equivalence, [1907])
- Gravity and Geometry (general relativity, GR, [1915])
- Quantum Mechanics + SR = quantum field theory (QED [1948]/QFT)

- EM + Weak interactions = "electro-weak" theory, EW, with symmetry breaking below some energy near 100 GeV: $SU(2)_L \times U(1)_Y \to U(1)_{EM}$ [1968 and Higgs boson 2012] [L = Left handed, Y = hypercharge, EM = vector potential A_μ or photon, and C = Color for the group, $SU(3)_C$].

- Standard model (SM): EW+strong QCD interactions.

- Beyond Standard Model (BSM) subgroups:

- $U(1) \subset SU(2) \subset SU(3) \subset SU(5) \subset SO(10) \subset E8$?. SO(10) is sometimes considered as a good candidate for a GUT.

- ToE?, SUSY?, GUTs? [Theory of Everything, supersymmetry, grand-unified-theories].

3.1. **Problems Facing Physics Beyond the Standard Model (BSM).** A list includes:
The problem of Dark Matter, the source of neutrino mass, why are there three generations of quarks and leptons, additional dimensions, the arrow of time, what is the mechanism for a bias of matter over anti-matter, how to calculate the 26 free parameters of the SM, a mechanism of inflation, and the mystery of Dark Energy along with constancy and small value of Λ when quantum field theory suggests a huge value.

My favorite conceptual problem with the Natural world is the incredible range of the Nature's variables and its ability to process those variables with mathematical precision over that range. There is a huge range of distance, time, energy, and particle masses. Energy of photons can range from Exa-electron-volts down to nano-electron-volts. And the biggest idea in the history of science is that the largest thing we can imagine was once smaller than the smallest thing we might imagine – our whole visible universe expanding from a tiny size smaller than a proton.

3.2. **Fundamental:** The word "fundamental" means foundational, a base support, un-derived, primal, essential, lying at the bottom or base of anything —ideally as "eternal truths." Physics Foundations refers to primary objects and theories, and theories are an end product of the "scientific method" with experimentally tested correctness and statistical assurance over time. Fundamental principles of a discipline of physics should aid models of physical events, should be expressed economically and efficiently — hopefully using a short list of mathematical statements from which other "lesser theories" can be derived or calculated. Ideally, the end concern of a quest for fundamentiality should be getting as close as possible to being able to find and state the foundational principles of a future "ultimate reality."

Some physicists suspect that current theories might be "effective field theories" from something deep down and different. But, if it involves territory anywhere near Planck units ($\sim 10^{-35}$ meters), it is unlikely to ever be testable. And many think that is where we do need to go for ideals like grand unifications. From a practical and human perspective, deepest concepts might be forever unattainable. It is best to believe that humans will attain a few deeper foundational levels beyond the standard model but still above the limit boundary.

It is hoped that future attainable collider energies will continue to reveal some deeper theories and that the idea of a "desert" below the standard model isn't true (i.e., between present day ~ 10 TeV to 10^{13} TeV or 10^{-18} to 10^{-31} meters [GUT level]). The limits to collider technology energy for the "imagined future" may be less than a PeV. The highest energy cosmic ray is an astonishing ~ 10 EeV (or 10^{10} GeV — compare to the expected Grand Unified energy near 10^{16} GeV or the Planck energy of 10^{19} GeV). Great GUT theories might end up partly "faith-based." So, the phrase "ultimate reality" is going too far, and we should restrict our discussions of "fundamental" to a depth that could actually be obtained. The "running" of key couplings for strong, weak and EM "forces" and also the "seesaw" mechanism for the puzzle of very light neutrino masses may point to a heavy GUT level particle — but we will never actually "see" it. There can be no assurance or trust in theories beyond experimentation.

4. QUANTUM MECHANICS, QM:

Copenhagen Postulates have been the standard textbook formalism of quantum mechanics since the 1930's — how the fundamentals are presented and used. The initial formalism of QM was "matrix mechanics" which was then largely pushed to the side by Schrödinger "wave mechanics" that enabled easier calculations and greater breadth (everyone knew wave mechanics, but almost no one knew matrices in the 1920's). The utility of matrices lives on in describing the quantum physics of angular momentum and group representations. But, a quite different interpretation of QM from Copenhagen is the "de Broglie-Bohm" 1952 Pilot Wave theory as a non-local, hidden variable theory without collapse! This minor interpretation still has strongly active supporters. There are now a vast variety of seemingly mutually exclusive formulations and interpretations.

Perhaps only one offers a visual glimpse of a possible mechanism in the quantum world: the 1986 transactional interpretation ("TI") of quantum mechanics from John Cramer (and newer version from Kastner) where psi is considered as an "offer-wave" to possible absorbers [3][12]. The otherwise mysterious Born rule postulate is "explained" by a "handshaking agreement" between the offer wave from an emitter and an advanced confirmation wave back from an absorber. Without the Born rule and completed transaction, we might ask, "What is the sound of one hand clapping?" In status, TI might still be improved and is not yet very popular. Another mechanistic (and heretical) goal may be to offer a "selection mechanism" for each particular chosen collapse of the wave function. Without that, we are forever stuck with the dogma of randomness and statistical probability and un-intuitive "probability amplitudes." Another great conceptual challenge is how "distance doesn't matter!" for entangled particles in "EPR" tests [Einstein, Podolsky, Rosen or "Bell" tests]. Explaining that might require a revolution in thought about the quantum world versus space-time.

Wittgenstein said the limits of my language are the limits of my world; and we lack words for the arena of possible mechanisms underlying the quantum world. Ruth Kastner encourages an acceptance of potential-reality as an extension of "reality"—"the underlying reality of possibilities" or "potentia." Lee Smolin is well-aware of present quantum mechanics as a bottleneck and says that we must "Resolve the problems in the foundations of quantum mechanics, either by making sense of the theory as it stands or by inventing a new theory that does make sense." [11]— and later adds, "This is probably the most serious problem facing modern science. It is just so hard that progress is very slow."

Depending on the hypercomplex algebra, there are various conjugations that can take numbers back to the reals [like hyper-conjugate or 'Hermitian Adjoint]." So, the "square root of reality" might be changed perhaps to a generic "star-root" symbolized by $\psi \simeq {}^*\sqrt{Prob} = P^{*/2}$ [10]. Physics has progressed by going down ever deeper into these "roots."
An example is an occasionally used representation of a single photon (the 'Riemann-Silberstein' form) found by taking the "square-root" (or "star-root") of its supposed energy density: $\psi^*\psi \propto (\epsilon_o/2)(E^2 + c^2B^2)$ becomes $\psi = \sqrt{\epsilon_o/2}\,(E \pm icB)$ [18]. The 'star root' operation is intended to only be heuristic and say "go to some higher dimension Clifford algebra." And we've mentioned Dirac theory as the square-root of the Klein-Gordon equation, $Dirac = (KG)^{*/2}$.

$$(1) \qquad \left[\partial^\mu\partial_\mu + \left(\frac{mc}{\hbar}\right)^2\right]\psi = 0 \qquad \xrightarrow{*/2} \qquad i\hbar\gamma^\mu\partial_\mu\psi - mc\psi = 0$$

where the γ^μ 's are the 4×4 unitary Dirac matrices.
The 'star-root' idea may go further, for example into the realm of quantum cosmology with supersymmetry where it is said that supergravity($N = 1$ SUGRA) naturally provides a Dirac-like 'square-root' of gravity [21]. And some programs for unifying general relativity and quantum mechanics use "tetrads" which can be thought of as the square-root of the metric, $g_{\mu\nu}$.

Quaternions and Pauli Matrices:

There is a well known story that William Rowan Hamilton had been walking along the Irish Royal Canal and suddenly realized his goal of an algebra that used three complex numbers. He immediately scratched them out on a stone bridge as: $i^2 = j^2 = k^2 = ijk = -1$ and later devoted much of his life to the development of this "quaternion" mathematics. It contained our dot-product and cross-product prior to the teaching of our "Gibbs-Heaviside" vector analysis and is 4D, so it was useful for special relativity and also for expressing Maxwell's equations. The Clifford symbol for quaternions is $C\ell(p,q) = C\ell(0,2)$ over real coefficients. The value in the p-slot stands for the number of real bases (meaning ${e_i}^2 = +1$). and the value in the "q" slot stands here for two "imaginary" bases (meaning ${e_j}^2 = -1$). Why just two?— because $i = jk = -kj$, so one base is covered by the product of the other two bases. The Pauli matrices should be considered as higher n than the quaternions, $\sigma_i = -iq_i$ or Pauli $\sim C \otimes H$ (bi-quaternions). For these matrices, we have $C\ell(3,0)$ with no imaginary bases. $C\ell(2)$ and $C\ell(3)$ are the $C\ell(n)$ short forms.

Quaternions excel in representing rotations in usual 3D space about some axis $\hat{n}$. A general quaternion is written as real coefficients q_i times bases e_i so that:

$$(2) \qquad q = q_0 + iq_1 + jq_2 + kq_3 = e^{i\cdot\hat{n}\theta/2} = cos(\theta/2) + i \cdot \hat{n}sin(\theta/2). \qquad q \cdot q = 1.$$

Notice that we automatically have half-angles so that a return to $\theta = 0$ requires rotating $\theta = 4\pi$ radians — twice around! Going once around, $\theta = 2\pi$ gives minus q. The real coefficients of a unit quaternion q plotted on i, j, k axes describe a unit 3-sphere or "hyper-sphere" S^3. Recall that the 3-sphere (angles χ, θ, ϕ) was once a favorite geometry for the case of a closed universe, but ours turned out to be flat Euclidean. For the usual sphere S^2, just set $q_3 = 0$; and for just a circle S^1, set $q_3 = q_2 = 0$ —it covers all three. The continuous (or Lie) group SU(2) has a Lie algebra $su(2)$ [the tangent space of SU(2)] generated by 3 elements – the quaternions! So hypercomplex numbers are also associated with our model groups.

Clifford Algebra:

In 1876, William Clifford invented new algebras to generalize Hamilton's quaternions for n dimensions and enhanced Grassmann's "wedge-product" algebra. Clifford algebra now applies to much of mathematical physics. In special relativity, SR, we have a metric $\eta_{\mu\nu}$ with signature $(+---)$ or $d(c\tau)^2 = +d(ct)^2 - dx^2 - dy^2 - dz^2$, a "quadratic form." For SR and Dirac algebra, we symbolize $C\ell(p,q) = C\ell(1,3)$ or just short form $C\ell(n=4)$ with dimension $n = p+q$ where q stands for the three minus-signs in the metric-quadratic-form. $C\ell(1,3)$ represents a large part of physics. To the untrained eye, actual computations using Clifford algebra may be unfamiliar and cumbersome; but there is a strong movement encouraging it for physics using the special case of David Hestenes' space-time or Geometric Algebra [15] with $C\ell(1,3)$. Most of us might just stay with the best familiar representation like Dirac matrices instead. The point is that Clifford algebra is a category encompassing most possibilities and may be relevant to future physics fundamentals. Let's propose that it will continue to apply to some higher "n".

Recall that our continuous or Lie groups are nested (see Section on Unification) so that lower groups are sub-groups "$\subset$" of the bigger higher groups. Clifford algebras dovetail with matrix representations of Lie Groups. Vector algebra is a sub-algebra of Pauli which in turn is a sub-algebra of Dirac. In 1961, Gell-Mann's $SU(3)_{flavor}$ "Eight-Fold-Way" used eight 3×3 "lambda" matrices – flavor at that time meaning just u, d, and s quarks. If we only used three of these instead, like $\lambda_1, \lambda_2, \lambda_3$ we would effectively get SU(2).

Isn't there a conflict between saying that the quantum world is "hypercomplex" and the need to satisfy humans? Well, some knowledge of quaternions (or Pauli matrices) is already needed for electron spin and to understand puzzles like electrons having to go twice around to return to the same state – a standard feature of quaternions. It is just a new learning curve. In practical applications, "Quaternions are now used throughout the aerospace industry for attitude control of aircraft and spacecraft [13]." And they are commonly used by "games programmers." There are many references showing how quaternions can be visualized [13] and computer programs that process Clifford algebras. But it will still be challenging to work out future heuristics.

5. Conclusions

It is proposed that the depths of different foundational theories of physics have levels quantified by the dimension, n, of their Clifford algebras $C\ell(n)$. Underlying the mechanisms of the quantum world is "sub-real" physics somehow mapped to hypercompex numbers that can enter "reality" by "Born-rule" squaring, $\psi^*\psi$ of presently unknown cause.

To go further, it is a human necessity to find a way to understand this and make it seem reasonable. Grasping a deepest "ultimate reality" is probably blocked by the limitations of our ability to conduct relevant experiments. Seeing beyond this barrier is a theoretical exercise of human intelligence without assured foundation or testability. That probably includes the whole field of "quantum gravity."

References

[1] Walter Isaacson, Einstein, His Life and Universe, Simon and Schuster, 2007 [pgs. 15, 335].

[2] John Baez,"How many fundamental constants are there?," http://math.ucr.edu/home/baez/constants.html

[3] Ruth Kastner, Understanding Our Unseen Reality:Solving Quantum Riddles, Imperial College Press, 2015.

[4] Zoltan Fodor, Christian Hoelbling,"Light Hadron Masses from Lattice QCD," https://arxiv.org/pdf/ 1203.4789, Mar 21, 2012

[5] Mark S. Gordon, Ed., Fragmentation: Toward Accurate Calculations on Complex Molecular Systems. Wiley, 2017, ISBN: 978-1-119-12924-0

[6] P. W. Anderson,"More Is Different," Science, New Series, Vol. 177, No. 4047. (Aug. 4, 1972), pp. 393-396.

[7] G. Weinstein, "Max Born, Albert Einstein and Hermann Minkowski's Space-Time Formalism of Special Relativity," http://ArXiv.org/pdf/1210.6929

[8] Steven Weinberg, Facing Up, Science and its Cultural Adversaries, Harvard University Press, 2001 [p. 185].

[9] Mary Sanders Pollock, Storytelling Apes Primatology Narratives Past and Future, 2015, The Pennsylvania State University Press,

[10] David Lyle Peterson,"Physics Lives in Form Heaven," FQXi Trick or Truth Essay Contest 2015, http://fqxi.org/community/forum/topic/2351

[11] Lee Smolin, The Trouble with Physics— the rise of string theory, the fall of a science, and what comes next, Houghton Mifflin, 2006

[12] Cramer, John G. (July 1986). "The Transactional Interpretation of Quantum Mechanics". Reviews of Modern Physics. 58 (3): 647-688. Bibcode:1986RvMP...58..647C. doi:10.1103/RevModPhys.58.647.

[13] Andrew J. Hanson, Visualizing Quaternions, Series in Interactive 3D Technology, 498 pages, Morgan Kaufman Publishers (Elsevier) 1st Edition, 2006.

[14] Sunny Y. Auyang, How is Quantum Field Theory Possible?, Oxford University Press, 1995.

[15] David Hestenes, Space-Time Algebra, Springer International Publishing, 1966, 2nd Ed. 2015

[16] Clifford Algebras and their Applications in Mathematical Physics, edited by A. Micali, R. Boudet, J. Helmstetter, Springer, 1992.

[17] Theodore Frankel, The Geometry of Physics, An Introduction, Second Ed., Cambridge, 1997, 2004, p. 502, Section 19.2d.

[18] Iwo Bialynicki-Birula and Zofia Bialynicka-Birula, "The role of the Riemann-Silberstein vector in classical and quantum theories of electromagnetism," 28 Dec 2012, http://arxiv.org/pdf/1211.2655

[19] Dave Peterson, "Learning Quantum Mechanics and Relativity," sackett.net/LearningQuantumMechanics.pdf, 2015.

[20] Andrew Liddle, An Introduction to Modern Cosmology, 2nd Ed., Wiley, 2003, 2008.

[21] Paulo Vargas Moniz, Quantum Cosmology - The Supersymmetric Perspective - Vol. 2: Advanced Topics, Springer, 2010.

[22] Frank Wilczek, The Lightness of Being, Mass, Ether, and the Unification of Forces, Basic Books, 2008.

[23] M. D. Maia, "Geometric Phase in Quaternionic Quantum Mechanics," 2001, https://arxiv.org/pdf/hep-th/0107107.pdf

[24] Quantum Field Theory, Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/entries/quantum-field-theory/ 2006,2012

[25] Bruce A. Schumm, Deep Down Things, The Breathtaking Beauty of Particle Physics, Johns Hopkins University Press, 2004

6. APPENDIX

Comments from other FQXi Essay authors:

Complex and hypercomplex numbers are parsimonious as needed: an example is quaternions H can also be represented by 2×2 matrices with complex entries or 4×4 matrices with real elements. [Reply to Ekhard Blumschein].
"Spacetime algebra $C\ell(1,3)$ naturally includes the spin group $Spin(1,3)$ leading to Fermions which are spinors." [Cristi Stoica, Indra?s Net]. Color has $C\ell(\bar{3} \oplus 3)$ having Dirac as a subalgebra.

Matrix representations of Lie groups are associative. So Octonian algebra cannot be directly used since it is generally non-associative for multiplication [Rick Lockyer, “Truth”].

Present day quantum mechanics is ostensibly a theory about the results of measurements. A final theory should not be “about the acquisition of knowledge.” [T. Durham].
Bell: the result of a measurement does not in general reveal some pre-existing property of the system, but is a product of both system and apparatus. Zeilinger: the results of observation are not always given prior to and independent of observation.
There is a good reason why Chemistry is not called molecular physic– almost nothing can be derived – it is an autonomous science and its basic principles are, for all practical purposes, fundamental. [Marc Seguin]

“Rather than being skeptical, we can try to be optimistic that the theories may eventually be indirectly testable, potentially yielding some novel predictions in regimes that are accessible to us. Similarly, we may be able to put more experimental constraints on the theories from the other direction, using observations in currently accessible regimes.” [Karen Crowther, 1/20/18 FQXi].

Photons and Light

Dave Peterson, 12/17/18 –3/6/19 [V-1], 4//18/19 [Version 2].

Abstract:

Although the concept of photons was suggested by Einstein in 1905 and generally accepted after the Compton effect of 1923, convincing evidence that light actually traveled as photon quanta had to await later experiments such as the photon "antibunching" effect of 1977. That essentially means that more than fifty years of the history of photons usually presented to physics students has been misleading. There exist more "neo-classical" explanations against "the three nails in the coffin of the wave theory of light" (black body radiation, photoelectric effect and the early Compton effect). And that has continued through the Einstein A, B coefficients (stimulated laser and spontaneous emission), approximate Lamb effect, and Casimir effect. But, precision tests of high order Feynman diagram calculations do show agreement with quantum electrodynamics over neo-classical approximations. And there are now many experiments using just a few photons at a time and single photons where quantum electrodynamics clearly applies and classical electromagnetism does not. Mostly, we can say that photons definitely exist; however, a strongly consensus definition of the words "photon" and even "existence" is still illusive. Some of the debate still depends on one's "interpretation of quantum mechanics."

INTRODUCTION:

For humans, light is defined as something that can stimulate the visual receptors in the eye thus enabling vision. It is electromagnetic radiation that travels at speed "c" in free space and can have an awesome range of wavelengths and frequencies, $c = \lambda f$. In addition to having an intensity, transverse polarization, and direction, it transfers quanta of energy to absorbers at single locations as would be expected for "particles." This "wave-particle" duality has never been consistently nor precisely defined and also applies to electrons and other quanta. A central connundrum has been whether light travels as photons (or superpositions of photons) – and this is not yet universally resolved.

Do photons exist? -- and how can they be defined? Quantum electrodynamics ("QED") says that electromagnetic radiation is quantized, and many arenas of modern quantum optics assume the existence of photons. But to what degree has that actually been verified experimentally when compared to "more semi-classical theories where the electromagnetic field remains classical?" How far can semi- or neo-classical physics suffice in place of photons as quanta of the electromagnetic field? Are "virtual photons" real? A major attribute of photons as particles is their degree of localization in space and time. The cases where photons are strongly claimed are examples where space-time localization was "clear" (for example like having a regular stream of single photons or the many current experiments with single or entangled photons [e.g., Pete19]).

Unlike mathematics with its clearly stated definitions, basic objects in modern physics have definitions that evolve with newer experimental discoveries and theoretical models. Sufficient definitions for the term "photon" or "electron" are rarely offered in textbooks. And if they were, there would be disagreements about them among physicists; and some disagreements would depend on a favored interpretation of quantum mechanics.

We might say that a photon is a relativistic, massless, charge-less particle of energy $E=\hbar\omega$ also corresponding to electromagnetic waves. And we add that it has helicity ± 1 -- making it a boson as an excitation or quanta of an underlying electromagnetic photon field. But a single photon might be prepared in a mixture of spin states or superposition of QM number states instead of just one. Alternatively, it can be said to possess transverse polarizations orthogonal to its momentum [direction of $p = h/\lambda = \hbar k = \hbar\omega/c$ = energy/c]. A newer attribute in our 21st century is that photons can also possess an orbital angular momentum ℓ of any number of $\hbar$'s – interleaved helical waves with no amplitude in their centers.

QED often states photon fields using the 4-vector potential, A^{μ} ($A^{o}=\phi$; A^{1}, A^{2}, A^{3}) Instead of E and B fields which themselves are derivatives or curls" of the A field. Photons may behave like electromagnetic waves possibly in some sort of wave function that is in some sense also supposed to be a "particle" that is different from any known classical particle. A particle aspect at least requires some degree of space-time localization so that we can say that something is "here" within some volume. But, Imagining a wave-function for a photon is hard because it lacks a "position state" $|r\rangle$ -- "there is no general particle creation operator that creates a photon at an exact point in space [Scully]" – a QED photon is inherently delocalized.

So, defining a photon is difficult except for a special case: **"A photon is what a photodetector detects"** [Roy Glauber, 2005 Nobel Prize]; and "a photon is where the photodetector detects it" via absorption [Scully]. Experimentally a photon is a "synonym for a discrete event, clicks of a detector or appearance of spots on a photographic plate [Rash]. Richard Muller says, "the photon is an event, not a thing." It does not exist until it is detected, then vanishes. So at least detection of photons at a local point can be discussed.

But our early physics educational system seems to present photons as objects traveling through space-time and having continuing existence between their initial creation and final measurement. Can that view be justified? Stating what is real between observations breaks a quantum rule that "unobserved things have no properties whatsoever." Some physicists state that photons as particle quanta do not exist but are instead just manifestations of the interaction of light with matter (e.g., Willis Lamb [Nobel prize 1955], Alfred Lande, E.T. Jaynes, [Rash]). Lamb wrote "…there is no such thing as a photon. Only a comedy of errors and historical accidents led to its popularity among physicists and optical scientists."
Let's take a moment to examine beliefs like this:

Students are initially presented with three historic reasons for rejection of classical waves in favor of quantized photon radiation ("the three nails" in the coffin of the wave theory of light): the blackbody spectrum, the photo-electric effect, and the early Compton effect. But the arguments against classical wave theory for these traditional cases are somewhat flimsy, and the topics can be covered with theories lacking "fuzzy ball" photons. An early theoretical approach to these cases was called "semi-classical" ("SC" = EM + QM) and treated the electromagnetic field in the usual classical sense with quantization only being present in atoms and molecules that can create or absorb these waves. Schrodinger was an advocate and originator of this view. He disliked the term "probability amplitude" for wave functions and believed that they instead represented "real" waves. For bound electron systems, $\psi^*\psi$ in reality represented an electron charge density $\rho = e|\psi|^2$. Historically, "Copenhagen" positivism opposed this view and its Born-probability idea quickly won out. But realist wave discussions continued on the sidelines for many decades and had some unexpectedly significant successes.

Historically, Einstein's 1905 "light-quantum hypothesis was consistently **rejected** by the physics community" until publication of the Compton effect experiment of 1923 ($\mathbf{p}=\hbar\mathbf{k}$ is implied for photons). Prior to that, radiation such as Thomson scattering was interpreted in a semiclassical way. Also, it was later realized that the "essence of the photoelectric effect does not require the quantization of the radiation field [Scully]." And, at room temperature, the "work function" Φ in the photoelectric formula $K_e=hf-\Phi$ is very poorly defined; and the idea of a stopping potential is physically impossible [Klassen].

Black body radiation is of course consistent with the idea of energy exchanges $\Delta E=n\hbar\omega$ between atoms and fields, but the quantization of radiation itself is not required. It was also realized that the Compton effect to lowest order was not real proof, and even the famous Lamb-shift of 1947 was challenged. And then, in addition, Einstein's stimulated emission and absorption "B-coefficients" of 1916 [e.g., Laser theory] can also be described by semi-classical theory in which the atomic electron cloud $\psi^*\psi$ acts like an oscillating dipole charge density. Laser light "photon number statistics" itself is also described semi-classically. That is, lasers "produce light beams that are in Glauber coherent states" whose photodetection analyses agree with those of the semiclassical theory.

Reasons for believing in photons are more subtle and often lie in experiments with just a few photons and in the improved accuracy and applicability of QED over SC (eg., the 1928 Klein-Nishima formula from QED is an improvement on the simpler early Compton effect, and some measured spectral line widths may be better derived from QED). Currently all existing experimental evidence agrees with QED. This has never failed while semi-classical theory often differs from experimental results

QUANTUM OPTICS:

The introduction to a popular quantum optics text says [Fox, 2006]:
"Quantum optics is the subject that deals with optical phenomena that can only be explained by treating light as a stream of photons rather than as electromagnetic waves." Below this level and prior to 1980 lay semi-classical theory that happened to be **"quite adequate for most purposes."** That is, until fairly recently, there were few topics that could not "be explained in the semi-classical approach." These supposedly included spontaneous emission and the Lamb shift and now photon antibunching. Contrary to traditional teaching, this does not include the photoelectric effect of 1905 because it can "be understood by treating only the atoms as quantized objects, and the light as a classical electromagnetic wave." That is, Planck's initial reaction was justifiable that the photoelectric effect and black body radiation only showed that "something is quantized" and not necessarily light itself.

"Sub-Poisson" light statistics is a relatively new topic in quantum optics that requires photonic light. Ordinary coherent laser light with constant intensity instead has photon number statistics that obey the Poisson distribution: $P(n) = n_{ave}^{\,n} \exp[-n_{ave}]/n!$ for n = 0,1,2, integer values. This has a standard deviation $\sigma_n = \Delta n = (n_{ave})^{1/2}$. Note again that a description of coherent laser light can be achieved using only semiclassical theory [Sudarshan, 1963]. Ordinary classical light with some time-varying intensity lies in the category of super-Poisson statistics so that $\Delta n > (n_{ave})^{½}$. Sub-Poisson statistics has a tighter distribution than these, and an example is a regular series of photon pulses (with no standard deviation, Δn=0 – quantum computing prefers this). As example, this case is seen in "photon antibunching" experiments where "a driven atom is unable to emit two-photons at once" {Kimble, Mandel, Dagenais, 1977} – two detections near zero separation do not occur. So what about Planck black body thermal radiation that is supposed to require discrete light? It has a variance of $(\Delta n)^2 = n_{ave} + n_{ave}^2$ which is greater than just n_{ave}. So it is super-Poissonian with a semi-"classical interpretation in terms of fluctuations in the light intensity." That is, again, something is quantized, but not necessarily the light. One reason sub-Poisson light is a new topic is that it is very hard to observe and depends on highly efficient detectors (which are becoming available). Any noise sources or detector imperfections introduce randomness into the light stream degrading its detection to just normal random light.

A key concept in quantum optics and laser theory is the **"Coherent state"** (symbolized by $|\alpha\rangle = |\ |\alpha|\, e^{i\phi}\rangle \in \mathbb{C}$) as the QM equivalent of classical electromagnetic waves [Glauber]. Light is a wave, and all wave phenomena can be related to harmonic oscillators. "**Number states" |n** ⟩ have energies $H|n\rangle = (n+ ½)\, \hbar\omega\, |n\rangle$. The coherent state is DEFINED by amplitude $|alpha\rangle = |\alpha\rangle = \exp(-½|\alpha|^2) \cdot \Sigma_o\, \alpha^n\, |n\rangle/\sqrt{\ }\, n!$ (which intuitively looks like a "square root" of a Poisson Distribution – as might be expected for probability amplitudes {"living in the square-root of reality"}).

$|\alpha\rangle$ is the right eigenstate of the annihilation operator a, $a|\alpha\rangle = \alpha\,|\alpha\rangle$ [FOX p. 158-159] $\langle n|\alpha\rangle = \exp(-½\,|\alpha|)\, \alpha^{n}/\sqrt{\ }n!$,

Then, apply the Born rule, $p(n) = |\langle n|\alpha\rangle|^2 = |\alpha^2|^n \exp(-|\alpha|^2)/n! = \eta^n e^{-\eta}/n!$ Where $\eta \equiv$ average n = <n>. And $|\alpha|^2 = \eta$. The result is a Poisson distribution for p(n), and η is a rate. Poisson processes or streams have the desirable property that their mergings or branchings are also Poisson distributions. Alternatively, if interarrival times between events is an exponential distribution, then mergings or branchings of streams also have exponential interarrival times.

"CAVITY QUANTUM ELECTRODYNAMICS" (CQED):

CQED "is the study of the interaction between light confined in a reflective cavity and atoms under conditions where the quantum nature of light photons is significant."

In free space, two level atoms in an excited state undergo spontaneous emission to the lower state with a characteristic lifetime (say nanoseconds). But an atom placed in a small reflective cavity can have a very different decay lifetime that can be strongly enhanced or even totally suppressed. This is an example of the "Purcell effect" of 1946 that was finally demonstrated at optical frequencies in 1987. Decay is not intrinsic to an atom but also depends on its environment: "the transition rate for the atom-vacuum (or atom-cavity) system is proportional to the density of final states" (this is called Fermi's Golden Rule).

Testing "Cavity Quantum Electrodynamics" (CQED) at optical frequencies is really hard and requires sub-micron sized cavities. Serge Haroche (Nobel Prize 2012) was a prime contributor to this new field and did early experiments using "Rydberg" atoms with decay photons in the micro-wave range (1 meter to ~1 mm wavelengths) – much easier. These are atoms such as in a cesium beam having very high quantum numbers (like n = 40) and hence very large outer orbits and very weak decay transitions. He explains in simple language: "In classical terms, the outermost electron in an excited atom is the equivalent of a small antenna oscillating at frequencies corresponding to the energy of transition to less excited states, and the photon is simply the antenna's radiated field. When an atom absorbs light and jumps to a higher level, it acts as a receiving antenna instead" [Haroche_SA]. A highly reflecting cavity only allows wavelengths that fit (λ/2, 1λ, 3λ/2, …) If an atom is placed in a small cavity having a size smaller than the transition λ/2, no photons can propagate and the atom is unable to decay at all. And for low frequency photons, "there are no vacuum fluctuations to stimulate its emission by oscillating in phase with it." But for precise sizes that fit the wavelengths, emission can be substantially enhanced. His work also "led to the creation of new kinds of microscopic masers that operate with a single atom and a few photons or with photons emitted in pairs in a two-photon transition" [Haroche_PT]. He also experimentally proved the principle of quantum decoherence. More optically oriented tests used cesium beam atoms with only principle quantum number n = 5 or 6 and wavelengths of 3.5 μm in tiny 2.2 μm cutoff cavities.

The Einstein B coefficients for stimulated emission or absorption is a weakly driven case of a more general phenomena that includes "Rabi oscillations." Instead of an electron simply jumping up or down a level, the electron can oscillate back and forth between the two levels at the Rabi angular frequency, Ω_R" [Fox]. It is difficult to observe this frequency because the radiative lifetime (e.g., from spontaneous emission) has to be longer than the oscillation period; and that requires high laser power.

The usual Einstein A coefficient for spontaneous emission is derived from semi-classical physics for atoms in free space (or in a large cavity). This A-rate decay time is altered when in small cavities. CQED physics is needed for the case of "strong-coupling" between atom and cavity and is described by the Jaynes-Cummings" model of 1963. The two-level atom interacts with "a single mode of the radiation field" – quantized light with small photon numbers.

SEMICLASSICAL VS QED

Quantum electrodynamics is considered as the most effective physical theory in history with no experimentally tested exceptions in its realm of applicability. Since it works so successfully, why bother to consider weaker theories with more classical aspects. One reason is that they seem to suffice more often than previously expected and sometimes offer adequately useful levels of approximation to the more accurate QED ["FAPP" —"for all practical purposes"]. They also correct the historical biases of the development of quantum theory. Another reason is that partly-classical theories are simpler and easier to understand while QED calculations may be hard, opaque and difficult to interpret. Calculations of QED perturbations led to numerous infinities and lacked a firm foundation (e.g., a dovetailing with axiomatic quantum field theory, AQFT). The rules of QED are more a set of practical algorithms for computation. The issue is how far we can take "weaker" theories before QED is strictly required. In pursuing this, we learn more about the essential features of the relevant physics – in particular the importance of the "zero-point-fluctuations" (ZPF of $\hbar\omega/2$ minimal energy) inherent in QED. Something resembling this <u>has to be added</u> to the older semi-classical theory to make them more effective. Although semi-classical theories (SC's) can explain the laser (the Einstein "B" coefficients), the Einstein <u>spontaneous</u> "A" coefficients of 1916 and the Lamb shift seemed to require something like ZPF. While a classical ZPF could explain most of the Lamb shift, there remains a three percent strictly QED contribution that is due to the polarization of the vacuum from virtual electron-positron fluctuations.

NEO-CLASSICAL THEORY (NC):

The next level of experimental discoveries above the quantum “three nails” were more of a challenge to semi-classical views and required the **addition** of new concepts resulting in what is called “neo-classical theory” --“NC,” e.g., [Jaynes]. This new approach required the inclusion of a **radiation reaction** or “back” reaction field on top of classical electrodynamics and the Schrodinger equation (EM+SEqn+BR). As a challenge to the prevailing view, Edwin Jaynes wished to consider more classicality to QED calculations and develop a better description of two-level atoms interacting with a quantized mode of an optical cavity (quasi-classical cavity dynamics, 1963 [Jaynes]). His modification allowed an atom to react back on an applied field with a radiation reaction that can lead to radiative damping and was able to calculate the Einstein A and B coefficients for spontaneous and stimulated emissions. This “neoclassical” theory is the most successful semiclassical theory for explaining spontaneous emission. Jaynes stated that his neo-classical NC theory “reproduces almost quantitatively the same laws of energy exchange and coherence properties as the quantized field theory, even in the limit of one or a few quanta in the field mode.” (!) A variant of this newer neo-classical theory is “random electrodynamics” also called background stochastic electrodynamics (SED) incorporating background Lorentz invariant random classical electromagnetic radiation. {Note: how can this be done? A zero-point-spectrum can be independent of an observer’s speed because of compensating changes in frequency and intensity. When an observer is approaching a radiation source, all frequencies will be shifted to higher values and all intensities are increased just so. And then, moving away from the source will have the opposite effect }.

With this, one can derive van der Waals forces and the Casimir effect [Boyer] – the attraction in vacuum of two parallel conducting plates with separations of microns. So, even classically, the vacuum can be viewed as not empty but buzzing with weak electromagnetic waves. But, again, the concept of radiation reactions also works equally well.

QED (quantum electrodynamics) uses creation and annihilation operators such as “a-dagger,” $a^+|n\rangle = |n+1\rangle\sqrt{}(n+1)$, that raises a state number count by one more photon. [For matrices, “dagger” also means “adjoint” or conjugate-transpose or “Hermitian” conjugate]. These are taken from and are similar in appearance to the raising and lowering operators for the quantum harmonic oscillator energy levels but are now used for number of particles instead of energy level numbers. To interpret a^+ in more familiar terms: “when there are n other identical Bose particles present, the probability that one more particle will enter the same state is enhanced by the factor (n+1)” [Feynman III]. “The presence of the other particles increases the probability of getting one more.” The square-root $\sqrt{}(n+1)$ is used because quantum mechanics works with “probability amplitudes” prior to quantum probabilities.

In QED, calculations using “normal ordering” of these operators (creation operators kept to the left of annihilation operators, “a^+a ”) removes the ZPF so that the entire remaining contribution to radiative frequency shifting comes from the radiation reaction. But, when the opposite “anti-normal” ordering aa^+ is used instead, vacuum field fluctuations become the cause (so we have a matter of interpretation). Vacuum fluctuations can be considered a physical basis for radiative frequency shift, but radiation reaction is an equally valid basis at this level of approximation. One might say that spontaneous emission and the Lamb shift are consequences of radiation reaction. And sometimes, NC is using ZPF's in disguise. But, with more generality, both classical ZPF and radiation reaction could be included. The equivalence of these two points of view is called the "fluctuation-dissipation theorem" [Mil_spon].

ADVANTAGES OF QED DERIVATIONS

The examples for the advantages of QED quantized fields above and beyond that of the semi-classical theories include:

Many interactions have higher order perturbation Feynman diagrams yielding more elaborate QED cross sections that agree with precise experimental measurements. SC's may only agree with low order calculations. The word "order" refers to the number of vertices in a diagram. So, for example, e-e scattering using an intermediate virtual photon would have two vertices and be of order two. "The full power of the quantum field theory will be seen at higher orders" {Kaku}.

The concept of fluctuating zero-point fields of energy density $\hbar\omega/2$ per mode is an important prediction that came with the quantum theory of radiation. These vacuum fluctuations of QED have been used to explain spontaneous emission and the Lamb shift. Quantized fields are also needed to calculate the anomalous magnetic moment of the electron. Beyond that, many multiparticle entanglement experiments do require quantized electromagnetic fields (e.g., quantum beats, quantum erasure, second-order photon correlations, two-site down-conversion interferometry [Scully]. More recently, photon anti-correlation experiments have been considered as proof that light is made of particles.

A major problem confronting Schrodinger's older idea of a "real" wave function was that of multiple particle wavefunctions existing in configuration space instead of conventional space-time [e.g., n non-interacting disconnected particles are represented by a point in R^{3n} space – how "real" is that?]. Einstein-Podolsky-Rosen EPR entanglements also reveal a limitation to semiclassical radiation theories. Examples of EPR experiments include the Kocher-Cummins 1967 experiment with three level Ca atoms where two photons have orthogonal polarizations and are entangled with other and Clauser's choice of mercury Hg atoms in 1974. These were the first true single-photon tests. Semi- and neo-classical theory struggles with correlation effects in n-particle states while QED does not.

After cascade decay sources, non-linear crystal SPDC became the standard source of single photons (Spontaneous Parametric Down Conversion from an intense laser beam that produces two entangled photons -- one of which can be used to signal (herald) the existence of the other). But this technique has very low and random yield. Modern quantum computing desires a regular source of photons to process. At present, "the most common sources of single photons are single molecules, diamond color centers and quantum dots" {Wikipedia}. It is now possible to supply streams of identical photons on demand.

VIRTUAL PHOTONS

Virtual photons in Feynman diagrams represent electromagnetic "forces" between charged particles and are themselves undetectable and should not be considered "real." They violate conservation of energy/momentum in accordance with an approximate uncertainty rule, $\Delta E \Delta t > \hbar/2$. [e.g., $\Delta p \Delta x \sim \Delta(pc)\Delta(x/c) \sim \hbar$ where Δ means σ (stdev)]. This is called being off the "mass-shell."

A recent article said:

"According to the received view Feynman diagrams are a bookkeeping device in complex perturbative calculations. Thus, they do not provide a representation or model of the underlying physical process. This view is in apparent tension with scientific practice in high energy physics, which analyses its data in terms of "channels" – the "Feynman-Dyson split" [Passon]. Feynman (1949) believed they represented actual particle processes, but it was Dyson who derived them from proper mathematics and noted conflict with realistic interpretations. Prior to Dyson's publication, "Nobody but Dick could use his theory, because he was always invoking his intuition to make up the rules of the game as he went along…

"Incoming and outgoing lines represent asymptotically free states and correspond to Dirac spinors (fermions) or polarization vectors (photons) in the calculation.

A "real" photon is massless with $k^2=k^\mu k_\mu = 0$ and has only two polarization states, whereas a virtual one, being effectively massive, has three polarization states. Virtual particles are also viewed as excitations of the underlying fields, but appear only **as forces**, not as detectable particles. They are "temporary"

The absorption or emission of a 'real' photon by a free particle of nonvanishing mass violates conservation laws. In addition, the popular picture of a single vertex (billiard ball collisions) for the Compton effect $\gamma + e_{rest} \rightarrow \gamma' + e'$ is disallowed in QED calculations (there is no γγe 1st order Feynman diagram—at least 2nd order is needed (and then it changes name to "Klein-Nishima" scattering).

DISCUSSION

We know that quantum mechanics has severe problems with interpretation. But, quantum field theory is worse and more conceptually intangible (and again, how can one define the words "photon" and "electron" as "quanta"). Standard problems with QED include the lack of dovetailing of mathematically rigorous axiomatic AQFT to the usual practical or Lagrangian LQFT, and the divergence of the S-matrix series expansion and its non-rigorous renormalizations. One result from AQFT is the conundrum of Haag's theorem (1955): the inability to transform from a free field theory to one with interactions.

"Haag's theorem is very inconvenient; it means that the interaction picture exists only if there is no interaction" [Streater and Wightman]. The Fock number representation for a free field cannot carry over to interactions – the number operator N(k) is not preserved (not a constant of the motion). That is a problem for understanding quanta. A field is not made up of numbers $|n(k_1), n(k_2)\ldots n(k_i)\ldots\rangle$ in momentum modes; $n(k_i)$ "quanta show up in an appropriate measurement" – an end point [Auyang].

Intuitively, one might wish to say that "real" quantum properties are "visualizable;" but that rarely applies to amplitudes and complex quantum states (nor hypercomplex mathematical descriptions like quaternion spinors nor gamma-matrices {entering the realm of "Clifford" algebras}). A better criteria for unobservable characteristics of a well-developed theory is called physical "kickability" [Auyang]: "something is kickable if it can be kicked and kicks back." A nice example is the Aharanov-Bohm effect in two-slit electron interference. Increasing the magnetic field inside a tiny solenoid near the slits increases its exterior vector potential A field which in turn alters the locations of ensemble constructive phase interference (the peaks move – the effect is "physical").

On the other hand, "Eigenvalues are not properties of quantum objects." They are not definite nor "kickable" but rather result from some principle of purely random selection (the Born rule).

A usual assumption is that energy and momentum (classical terms) apply to localized photons in flight. But, another possibility is that density of wave-lengths in space or time (k or ω) represents "code" for detected momentum or energy that is de-coded, actualized or "realized" only at absorption.

APPENDIX:

HARMONIC OSCILLATORS;

The linear harmonic oscillator is one of the classic topics in non-relativistic quantum mechanics for the Schrodinger equation and has a "spring" potential energy: $V = kx^2/2$ with spring constant $k = m\omega^2$. It has the interesting property that its quantum energy states have constant spacings so that the n'th single particle state has energy $E_n=(n+ ½)\hbar\omega$ with a non-zero ground state $E_o=\hbar\omega/2$. As example, this applies to infrared spectroscopy of vibrational levels of diatomic molecules. There are raising and lowering operators, a^+ ("a-dagger") and a, that increase the energy levels by one unit or take away one unit. The n'th level is created from the zero'th "vacuum" level by n applications of a^+ : $|n\rangle = (a^+)^n|0\rangle /\sqrt{\ } n!$.

This concept is carried over into QFT for the electromagnetic field of bosons but with a different interpretation. State $|n\rangle$ is no longer for just a single particle but is now a field state having n particles present all having the same energy, $\hbar\omega$. An example of a general state for particles (with or without mass) is, $|1\ 1\rangle = |k_1,k_2\rangle$ meaning momentum $\hbar k_i$ and energy $\hbar\omega_i$ in each state over 1+1 particles. It may be created from $|0\rangle$ by raising operators $a^+(k_1)$ and $a^+(k_2)$. For massless photon bosons, $\hbar\omega =\hbar k/c$, so the energy is proportional to the momentum, and only one needs to be used. State

$|k_1,k_2\rangle =|k_2,k_1\rangle$ by boson interchange with no sign change. The Hamiltonian operator is $H = \int d^3k\ \omega_k [N(k)+ ½]$.

ZERO POINT FLUCTUATIONS

ZPF is supposed to refer to the state or energy of the vacuum at absolute zero temperature. Classical non-relativistic statistical mechanics would claim all random energy is thermal. If it really exists in QFT, it is believed to consist of an $\hbar/2$ weight on every normal mode. But,it cannot be said to be well characterized primarily because of the huge discrepancy between the background ZPF energy predicted by theory versus its near absence in real observations. It's main selling point is usually the Casimir effect – but there are difficult theories that can produce it without ZPF. Then there is the Lamb shift of 1058 MHz between the energies of the 1s and 2s levels of hydrogen that is claimed to be due to electromagnetic ZPF but with a small 3% contribution from the polarization of the vacuum from electron-positron fluctuations. The Lamb shift is mainly due to the fluctuation of E and B fields from the QED-ZPF for k values $k\in [\pi/a_o, mc/\hbar]$. Frequencies higher than those associated with the Bohr orbit jitters the orbit. The perturbing ZPF electric field $E^2\propto \hbar/2$.

The claim of NC is that there is a background classical zero point fluctuating field independent of QED.

There is one infinite collection of harmonic oscillators we call the background "photon field", another we call the "electron field", and so on. Since a photon has no independent existence from the photon field of which it is an excitation, the photon is a derived concept. You start out with just the idea of a field. You model the field as an infinite collection of harmonic oscillators, … The electron, in a quantum mechanical description, is truly a "field" that permeates all of space, and the "excitation" is the region where the probability is highest. This is the basis of the QFT's description of a particle. Excitation refers to the amount of energy needed, to take a field from the vacuum state (ground state, energy zero) to an "excited state, corresponding to the creation of one particle. During the scattering process itself, the electron loses its recognizable individuallity, and all there is (meaningfully) is the quantum field with an indefinite particle content. QED gives the full observable answer, with fully dressed, free electrons entering and leaving the scattering event for t --> $\pm\infty$, but complex quantum field behavior at finite times.

The classical Hamiltonian for the electromagnetic field can be expressed as a continuous superposition over harmonic oscillator Hamiltonians: Classical uses a*, quantum uses "a-dagger:" $H = \int d^3k\Sigma_\sigma \hbar\omega(k)(a^+_{k,\sigma}\, a_{k\sigma})$ excited stationary states of the quantum EM field, which we will interpret as states with one or more photons. {http://www.physics.usu.edu/ /3700_Spring_2015/What_is_a_photon.pdf }. The behavior of charged particles can be affected by EM phenomena, even when no photons are present! This is the idea behind the "Lamb shift" found in the spectra of atoms. And it is the key idea needed to explain spontaneous emission of photons from atoms. Quantum fields are the stuff out of which everything is made!

The "electric" $E(^-)$ and $E(^+)$ operators

… have "the property of raising an n-photon state to an n+1 photon state" or lowering from n to n-1 state. The state $E(^-)(rt)|\text{vacuum}\rangle$ is a new one-photon state [Rash]. These operators should be assumed to use the previous annihilation/creation operators (but notice the sign conventions here which will pertain to - and + complex frequencies—clockwise or counter-clockwise in the complex plane). The Born rule for light still applies as probability $p \propto E^*E = |E|^2$, the intensity of classical light wave.
"Experiments which detect photons ordinarily do so by absorbing them," so detection processes represent photon annihilation using the complex field $E(^+)(rt)$ [Glauber63].
It is well known that the electromagnetic field may be treated as an assembly of harmonic oscilators (e.g., Schrodinger, 1926).

<u>We work in the complex plane</u> by first splitting the expression for an oscillating electric field E into two complex conjugate terms $E = E^+ + E^-$ where $E^- = (E^+)^*$ and E^+ has only positive frequency terms (i.e., those varying as $e^{-i\omega t}$. EE's are familiar with this as a mathematical convenience—like rotating clockwise ($e^{-i\omega t}$) versus CCW ($e^{+i\omega t}$) in the complex plane. Instead of x and p as operators, we now have quantization of <u>fields</u>.

Define a free transverse electric field **operator** as:

$$E^+(r,t) = i\,\Sigma(\hbar\omega_k/2\epsilon)^{1/2}\, a_k\, u_{k,\lambda}(r)\, \exp(-i\omega_k t) \qquad \text{[Orszag].}$$

{think of E as energy density amplitudes or electric field density and remember than energy density of electric fields is proportional to $E^*E=|E|^2$. }

where wave-function $u = e^{\lambda} \exp(ik\cdot r)/\sqrt{\text{volume}}$, and e^{λ} is polarization ($\lambda = 1,2$) . "a" and a-dagger are now part of these field operators. The sum $E= E^{+}+E^{-}$ is also expressed as $E_{\perp}(r) = i \sum \mathcal{E} [a_i \epsilon_i e^{ik.r} - a_i^{+}\epsilon_i e^{-ik.r}]$ [Tann], where

$\mathcal{E}$ refers to the electric field in reciprocal space, $\mathcal{E}$ (k)] where E and $\mathcal{E}$ are radial spatial Fourier transforms,

$\mathcal{E}(k,t) = (2\pi)^{-3/2} \int d^3r\, E(r,t)e^{-ikr}$, and $E(r,t) = (2\pi)^{-3/2} \int d^3k\, \mathcal{E}(k,t)e^{+ikr}$.

[for example $1/4\pi r \xleftrightarrow{FT} (2\pi)^{-3/2}/ k^2$].

Detectors are usually in a ground state so that only energy absorption can occur. E^+ takes an initial state $|\psi_i\rangle$ to a final state $|\psi_f\rangle$ where i is usually higher in energy than f. The transition probability for this is $W_{if} =|\langle\psi_f| E^+ |\psi_i\rangle |^2$.
The E^+ component of the electric field is proportional to the annihilation operator of the field. This is stressed because detector states are usually ground so that only absorption takes place during photodetection. E^+ takes an initial wavefunction ψ_i to a final wave function ψ_f where state i is usually higher than state f—so it lowers its energy and gives it to the detector.

In other words, Heuristically, annihilation of a photon, operator a, releases positive energy ~amplitude E^+ at a detector associated with positive frequency (going with $-i\omega t$) while the creation of a photon, a^+ (dagger) introduces a negative frequency component and a loss of energy in creating the photon, E^- going with $-(-i\omega t)$. Think of these positive and negative frequencies in terms of $\varphi = (k\cdot r-\omega t)$ so that + goes with $e^{+i\varphi}$ and – goes with $e^{-i\varphi}$.
Average field intensity $I_i(r,t) = \sum|\langle \psi_f| E^+| \psi_i\rangle|^2 = \sum\langle \psi_i| E^-| \psi_f\rangle \langle \psi_f| E^+| \psi_i\rangle =$
$\langle \psi_i| E^- \cdot E^+| \psi_i\rangle$. This is "the probability of observing a photoionization in a detector between times t and t + dt." The sequence $E^- \cdot E^+$ is called normal ordering in the Heisenberg picture.

As an application, consider Franson Interferometry -- a "space-like method for determining time-time correlations of entangled photons" Each photon of an entangled pair (1-biphoton) travels through a single-photon rectangular Mach-Zehnder (MZ) type path for a straight through short path and a long rectangular path resulting in two different photon output times. At the beam-splitter entry to the MZ rectangle, we create two new partial amplitudes using the E- operator (we are treating these amplitudes very similarly to the usual electric amplitude photon wave functions at beam splitters).

ELECTRON WAVES

Schrodinger's idea of the real continuous classically Maxwell wave perspective (as functions rather than as operators) for actualized photons was also intended to be carried over to "real" de Broglie waves for electrons (when viewed from an appropriate frame of reference). The following abstract is an example: [RASH2 Abstract]:

"In this paper, I argue that we can avoid the paradoxes connected with the wave-particle duality if we consider some classical wave field—"an electron wave"—instead of electrons as the particles and consider the wave equations (Dirac, Klein–Gordon, Pauli and Schrödinger) as the field equations similar to Maxwell equations for the electromagnetic field.

That interpretation considered $|\psi|^2$ as a measure of the distribution of the electric charge of electron in space. Schrödinger considered it possible to abandon both quantum jumps and corpuscular representations and to consider the electron as a wave packet described by the wave equation. The "Born rule for light" [prob $\propto$ $|E|^2$] is a trivial consequence of the Schrödinger equation, occurring only for relatively short exposure times, whereas for long-term exposure it is necessary to use a more general nonlinear rule. I propose to consider the following perspective: **there are no electrons as particles,** but instead there is an electron wave, which is a real classical wave field, in the sense that the wave is continuous in space and time. From this perspective, the Dirac equation is the equation of the electron field, similar to Maxwell's equations for the classical electromagnetic field. This enables consideration of the electron wave as a classical continuous field that has an electric charge, continuously distributed in space with density ρ, internal angular momentum, continuously distributed in space with density s and not connected with the motion of the electron wave, and an internal magnetic moment, continuously distributed in space with density m and unconnected to the motion of the electric charges of the electron wave.

Note that expressions which are the basis of the considered explanation of the shift, were obtained within the purely wave representations of electromagnetic and electron waves without the use of such concepts as "photon" and "electron".

References:

[Milonni] P. W. Milonni, Semiclassical and quantum electrodynamical approaches in nonrelativistic radiation theory, Physics Reports (Section C of Physics Letters) 25, No. 1(1976)1—81. North-Holland Publishing Company, Or web at:
http://web.physics.ucsb.edu/~phys250/fall2009/Ref3_Milonni.pdf

[Mil_spon] P.W. Milonni, "Why Spontaneous emission?", Am. J. Phys. 52 (4), Aprol 1984. P 340-343.

[Scully] Ashok Muthukrishnan, Marlan O. **Scully**, and M. Suhail Zubairy, "**The concept of the Photon—revisited,**" OPN Trends, October 2003. This is a great summary article on quantum optics up to 2003.
https://pdfs.semanticscholar.org/7255/948988f5280323a83d83237fb138a8243e6a.pdf
[Previous: M. O. Scully, M. Sargent III, "The Concept of the Photon," Physics Today 25, No. 3, March 1972].

[Glauber05] Roy J. Glauber, Nobel Lecture, Dec 8, 2005, One Hundred Years of Light Quanta. https://www.mpi-hd.mpg.de/personalhomes/palffy/Files/glauber-lecture.pdf
[Glauber63] Roy J. **Glauber**, the quantum theory of optical coherence, Phys Rev 130 (6) June, 1963, p 2529. Available at:
https://www.aps.org/newsroom/pressreleases/upload/pr.pdf

[Jaynes] E.T. Jaynes ; F.W. Cummings, "Comparison of quantum and semiclassical radiation theories with application to the beam maser," Proceedings of the IEEE Volume: 51 Issue: 1 , 1963.

And, C.R. Stroud Jr. and E.T. Jaynes, "Long-Term Solutions in Semiclassical Radiation Theory," Phys. Rev. Al (1970) 106. Or: https://bayes.wustl.edu/etj/articles/long.term.sol.pdf Also see: M.D. Crisp and E.T. Jaynes, Phys. Rev. 179 (1969) 1253. And, E.T. Jaynes, Coherence and Quantum Optics, ed. L. Mandel and E. Wolf (Plenum, N.Y., 1973) p. 35.

[Fox] **Mark Fox, Quantum Optics An Introduction,** Oxford University Press – Master series in atomic, optical, and laser physics, 2006

[Orszag] Miguel Orszag, Quantum Optics, 2nd Ed. 1997, 2008, Springer.

[Tann] Claude Cohen-Tannoudji, Jacques Dupont-Roc, Gilbert Grynberg, **Photons and Atoms,** Introduction to Quantum Electrodynamics, 1984, John Wiley.

[Boyer] Timothy H. Boyer, Random electrodynamics: The theory of classical electrodynamics with classical electromagnetic zero-point radiation, Phys. Rev. D 11, 790, 1975 And "Understanding the Planck Blackbody Spectrum and Landau Diamagnetism within Classical Electromagnetism," https://arxiv.org/pdf/1602.00295.pdf, 2016.

[Rash] S.A. Rashkovskiy - 2015 Proc. Of SPIE Vol 9570 file:///Users/Dave/Downloads/Are%20there%20photons%20in%20fact%253F%20(3).pdf
https://spie.org/Publications/Proceedings/Paper/10.1117/12.2185010

[RASH2] Sergey A. Rashkovski, Quantum mechanics without quanta: 2. The nature of the electron https://slideheaven.com/quantum-mechanics-without-quanta-2-the-nature-of-the-electron.html Tomsk State University, 36 Lenina Avenue, Tomsk 634050, Russiae-mail: rash@ipmnet.ru123

[Haroche_PT]: Serge Haroche and Daniel Kleppner, "Cavity Quantum Electrodynamics," https://www.chem.purdue.edu/courses/chm374/Articles%20etc/Haroche89.pdf , Physics Today January 1989, Vol. 42(1): 24-30.

[Haroche_SA]: Serge Haroche, Jean-Michael Raimond, "Cavity Quantum Electrodynamics," Scientific American, April 1993, https://qudev.phys.ethz.ch/content/QSIT16/QSIT16_Super_L03_slides.pdf

[Hobson] Art Hobson, "There are no particles, there are only fields," https://arxiv.org/pdf/1204.4616, 2012

[Brooks] Rodney A. Brooks, Fields of Color: The theory that escaped Einstein, 2010.

[Gurel] Atilla Gurel, "Source Field Effects and Wave Function Collapse," https://arxiv.org/pdf/quant-ph/0307157.pdf

[Auyang] Sunny Y Auyang, How is Quantum Field Theory Possible?, Oxford, 1995. http://www.creatingtechnology.org/papers/field.pdf

[Sudarshan] E.C.G. Sudarshan, "Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams," Phys. Rev. Lett. 10 #7, p 277, 1963 [Note, Controversy, Sudarshan presented his work prior to Glauber but wasn't considered for the Nobel Prize in Physics for this nor for his "V-A" theory of "left-handedness"].

[Feynman] R. Feynman, R.B. Leighton, M. Sands, The Feynman Lectures on Physics, Addison Wesley, 1963 (Volumes I, II, and III).

[Passon] Oliver Passon, “On the interpretation of Feynman diagrams, or, did the LHC experiments observe$H \rightarrow \gamma\gamma$?,” https://arxiv.org/pdf/1901.02233.pdf , 8 Jan, 2019, for European Journal for Philosophy of Science

[Pete19] Dave Peterson, Appearances of Retrocausality in Entanglement Experiments, 11/2/18 -1/4/19, “Five classic journal articles on early photon entanglement experiments are sampled for discussion.”

NOTES

{https://en.wikipedia.org/wiki/Nobel_Prize_controversies#Physics } And reference [Sudarshan].

Sudarshan began working on quantum optics at the University of Rochester in 1960. Two years later, Glauber criticized the use of classical electromagnetic theory in explaining optical fields, which surprised Sudarshan because he believed the theory provided accurate explanations. Sudarshan subsequently wrote a paper expressing his ideas and sent a preprint to Glauber. Glauber informed Sudarshan of similar results and asked to be acknowledged in the latter's paper, while criticizing Sudarshan in his own paper. "Glauber criticized Sudarshan’s representation, but his own was unable to generate any of the typical quantum optics phenomena, hence he introduces what he calls a P-representation, which was Sudarshan’s representation by another name", wrote a physicist. "This representation, which had at first been scorned by Glauber, later becomes known as the Sudarshan–Glauber representation."

In 2007, Sudarshan told the Hindustan Times, "The 2005 Nobel prize for Physics was awarded for my work, but I wasn't the one to get it. Each one of the discoveries that this Nobel was given for work based on my research."[17] Sudarshan also commented on not being selected for the 1979 Nobel, "Steven Weinberg, Sheldon Glashow and Abdus Salam built on work I had done as a 26-year-old student. For another major topic: Sudarshan regarded the “V-A theory” as his finest work. The Sudarshan-Marshak (or V-A theory – vector minus axial-vector theory exposing the existence of intrinsic “left-handedness.”

“On-demand Semiconductor Source of Entangled Photons Which Simultaneously Has High Fidelity, Efficiency, and Indistinguishability,” Hui Wang, et al., https://arxiv.org/pdf/1903.06071.pdf & Phys. Rev. Lett. 122, 113602
Focus: “Entangled Photon Source Ticks All Boxes.”
“A quantum-dot-based device combines all of the attributes necessary for producing a reliable source of entangled photons for quantum information applications.”

Abstract: An outstanding goal in quantum optics and scalable photonic quantum technology is to develop a source that each time emits one and only one entangled photon pair with simultaneously high entanglement fidelity, extraction efficiency, and photon indistinguishability. By coherent two-photon excitation of a single InGaAs quantum dot coupled to a circular Bragg grating bull’s-eye cavity with a broadband high Purcell factor of up to 11.3, we generate entangled photon pairs with a state fidelity of 0.90(1), pair generation rate of 0.59(1), pair extraction efficiency of 0.62(6), and photon indistinguishability of 0.90(1) simultaneously. Our work will open up many applications in high-efficiency multiphoton experiments and solid-state quantum repeaters.

EXPLAINING S ORBITALS AND BONDING

DAVE PETERSON

ABSTRACT. The simplest covalent atomic bonds are the cases of H_2^+ and neutral diatomic hydrogen H_2 beginning with the overlap of two S-orbitals. Understanding that bond is aided by an understanding of S-orbitals. In the overlap region, the 'information wave' ψ 'realizes' an enhanced negative charge density source via the Born rule, $P = \psi^*\psi$, and this enhancement can result in chemical bonding. Any initial candidate wavefunction, ψ, gets altered by the $\psi^*\psi$ electron-enhancement of orbital overlap. Interpretations and precise details of explanations of bonding lack consensus. The discussion here suggests that this basic foundation of quantum physical chemistry is partly clear in a mathematical sense but very unclear in an intuitive sense. Textbooks stick with the math and generally avoid any intuitive explanations.

1. Chemical Bonding:

It is generally accepted that a covalent bond is achieved by an effective enhanced formation of negative charge between two atomic nuclei − a"redistribution of electron density to yield a build up in the interatomic midpoint region." But even in 2008, there was still controversy in the details leading to the covalent bond [1]. Despite a history of great experimental and computational success, "it is remarkable that the physical explanation of the origin of covalent bonding is still a subtle and contentious issue generating much discussion." So, the reason that chemistry texts are so vague about the nature of the covalent bond is that they are still unsure exactly how to interpret the bonding mechanism. One typical initial approach is MO-LCAO − a molecular orbital from a linear combination of atomic orbitals. And then the Born rule $\psi^*\psi$ enhances the effect in the overlap region. One interesting aspect of this is that partial charge accumulates there, $dQ = e\psi^*\psi\, dVol$. This is in contrast to physical measurements which require a discrete whole charge to be transferred, and $\psi^*\psi$ is the probability of an electron being intersected in the experiment. It suggests that there is an intermediate interpretation of the Born-rule $\psi^*\psi$ for the case of bound state reinforcing orbitals separate from measurement.

There are often different equivalent approaches and interpretations for quantum mechanical problems. Feynman [3] considered $H_2{}^+$ binding in terms of an electron exchange similar to the 'flip-flop' of an N-atom in an ammonia molecule (NH_3). There is a special new energy term emerging in a two-state base system related to a tunneling entity flipping

Date: May 13, 2012.

email: davepeterson137@gmail.com. Paper updated to March 6, 2014.

'back-and-forth' as a resonance. That is, the electron of ${H_2}^+$ might prefer to be near one or the other protons for a "double-well" system [10], and the electron can pass through a potential maximum in the middle. Exchange causes a splitting of energy levels with one state lying lower than the other [E_I high and E_{II} low]. Essentially, the electron kinetic energy (KE) near midpoint can become negative so that momentum p can be imaginary. There is then a reduced net energy or a binding energy for the possibility of an electron jumping from one proton to another. This 'exchange effect' idea was used by Yukawa to aid his understanding of nuclear binding.

The main opponent of the idea of electrostatic attraction for chemical covalent bonding is Klaus Ruedenberg (1962 to present) [2]. His position on ${H_2}^+$ is 'that orbital sharing lowers the variational kinetic energy pressure and that this is the essential cause of covalent bonding." His detailed variational calculations allow for contraction of the size of a 1S orbital by a free parameter α so that in equation (3) below we can have $e^{-r/a_o} \to e^{-\alpha r/a_o} \simeq e^{-1.238r/a_o}$. [1] (for neutral H_2, we might have $\alpha \simeq 1.19$). It is not clear why this parameter should be allowed to vary. Having a higher $\alpha > 1$ causes higher kinetic energy but also stronger (more negative) potential energy. A step after this promoted contraction is overlap causing charge delocalization and charge redistribution. The electron belongs to both nuclei which lowers the KE. There is orbital sharing, orbital contraction, and orbital polarization. This minority view is almost never discussed in undergraduate chemistry texts.

The case of neutral diatomic hydrogen H_2 with two electrons adds the presence of two identical particles obeying an exclusion principle. The molecular wavefunction has to have not only even or odd parity over space but also be antisymmetric for interchange of space and spin coordinates of the two electrons [4]. We need a zero net spin ground state (anti-parallel spins) and again even parity leading to electrons spending most of their time in-between the protons causing binding (-4.476 eV and separation 0.74$\mathring{A}$). In general the strength of chemical bonds is due to the accumulation of electron density in the bonding region [11]. [2] The up and down spin electrons form a '$1S\,\sigma$' bond between protons. A wavefunction for the symmetric case may look like:

$$\Psi_S(r_1, r_2) = \frac{1}{\sqrt{2}}[\phi_a(r_1)\phi_b(r_2) + \phi_b(r_1)\phi_a(r_2)] \tag{1}$$

and a minus sign is used for the antisymmetric case, Ψ_A.

Note that technically, this formula (1) says that the two atoms of a hydrogen molecule are **entangled**. The modern interest in entanglement is for long distance "spooky action," but this is a short distance example. It is also true that the two electrons of a helium atom are entangled (measurements cannot be made on one particle without affecting the other).

[1]The Bohr orbit is $a_o \simeq 0.53\,\mathring{A}-$ which in 'atomic units' is just called one 'bohr.' Likewise, the reference energy $E_h = 27.21\,eV = 2.626 MJ/mol$ is called a 'hartree.'

[2]With some uncertainty in the literature for the case of ${H_2}^+$ ion where bonding is weak, and cause is subject to debate.

Be aware that there are many interpretations of quantum mechanics. One aspect of QM concepts is "wave-particle" duality. Feynman was a 'particle person,' but many other physicists believe in a wave or field-only interpretation. The electron-field in quantum field theory represents electrons. The 1S orbital in the hydrogen atom might not just represent an electron but may actually **be** the electron. A perceived particle nature might not show itself until a measurement occurs. Asking what an electron is doing in an atom assumes that an electron actually exists there. As an example, the de Broglie-Bohm 'pilot-wave' interpretation of QM would say that indeed particles do exist and have well defined trajectories. But unlike a 'standard interpretation, an electron does not move if it is in a stationary-state like the 1S or 'σ-bond.' The associated lack of any kinetic energy is offset by a specially devised 'quantum potential' $\propto (-\hbar^2/2m|\psi|)\nabla^2|\psi|)$.

A high-school level explanation of the ${H_2}^+$ covalent bond could be the following: An electron in its lowest energy state is like an exponentially decaying 'cloud' surrounding a proton. Suppose that on a piece of paper there is placed a quarter to the left and another quarter to the right standing for two protons each having a 'cloud' of **four pennies** lying to the left, right, up, and down directions and representing 'electron amplitudes.' If the quarters approach each other so that two of the pennies overlap at the midway point, M, then there will be two pennies at M. Could this double weight cause the protons to have a net attraction? No; they still have a net repulsion. But, there is a basic rule of quantum mechanics that the "probability of finding an electron at some location" goes as the **square** of the amplitude so that the 2 pennies at M will count as $2^2 = 4-$ an enhancement of electron density there. Now there is enough negative charge density at mid location to cause a net attraction, and chemical bonding will occur. [3]

So, how much charge is that? The repulsion of two protons by the inverse square electric field would be balanced against a single charge of 1/4th e at a mid point. The new Born enhanced overlap gives 4 pennies at the midpoint with another 6 at other positions for a charge ratio of 4/10 electron charges. However, the plane sheet layout isn't quite right and really needs at least four more pennies each lying above and below each proton. Then the midpoint charge is 4/14ths e $\simeq 0.286\, e > 0.25-$ so we still see bonding, but barely. The ${H_2}^+$ case is one of the weakest of chemical bonds, and H_2 gives stronger chemical bonding.

Going one step further for planar ${H_2}^+$, the enhancement of pennies at the midpoint is $4 - 2 = 2$ extra pennies. Quantum mechanics also allows the base states an electron on the left proton (ℓ) and an electron on the right proton (r) to add together symmetrically or also to subtract (anti-symmetrically and giving 'anti-bonding'). Call these states I and II.

$$|II\rangle = \frac{1}{\sqrt{2}}(|\ell\rangle + |r\rangle\ ), \quad |I\rangle = \frac{1}{\sqrt{2}}(|\ell\rangle - |r\rangle\ ) \tag{2}$$

[3] So, is that really seen for H_2? Plots of electron density at the midpoint between the two protons show a value that is about 3.8 times stronger than the corresponding distances on the opposite or back side.

[4] State II has the positive overlap and lower energy, and state I has zero overlap at M. For the pennies case, that means that state II has an excess of two pennies (more negative charge there), and state I has a deficit (zero minus overlap two is minus two). If these correspond to changes in energy, A, then we can explain an energy splitting from the non-overlapping free state: $E_I = E_o + A$ and $E_{II} = E_o - A$. The state E_I can be negative and represent a net attraction and hence chemical bonding.

2. 'S' Orbitals

The first two purely radial integral-square normalized 1S and 2S states of an atom are given by [4]:

$$\psi_1(r) = \frac{1}{\sqrt{\pi}}\left(\frac{Z}{a_o}\right)^{3/2} e^{-Zr/a_o}, \qquad \psi_2(r) = \frac{1}{\sqrt{8\pi}}\left(\frac{Z}{a_o}\right)^{3/2}\left(1-\frac{Zr}{2a_o}\right)e^{-Zr/2a_o} \tag{3}$$

where $a_o = 4\pi\epsilon_o\hbar^2/me^2$ is the first Bohr orbit $\simeq 0.53\mathring{A}$, and proton number Z = 1. The reduced electron mass should really be used $m_r = m_e/(1+m_e/M_p)$ so that $a_o' = a_o(1+m/M)$. These orbitals are solutions of the Schrödinger equation (SE) for an electron in a three-dimensional Coulomb field. And then there is also multiplication by a time varying with a frequency given by $\nu = h/E$. For the first $\psi_1(r,t)$, this is like a central pole circus tent shape that is up and then becomes inverted down and then back to up again. One initial curiosity is that exponential tails go out to infinity, but can the whole wave function change so fast that the tails are causally disconnected (beyond the speed of light). Not really, because c time a half wave period is about 460 angstroms which is out there pretty far. However, some view the wave-function as holistic with special quantum network type communication between all of its portions. This communication can be a-temporal involving both back and forth in time transmission effectively instantaneously so that far-flung portions work together well.

The ground state '1S' waveform solution can be most easily understood by simply 'assuming' an exponentially decaying profile: $\psi_1 = Ae^{-br}$ and plugging that into the SE: $-(\hbar^2/2m)\nabla^2\psi = (E-V)\psi$ to obtain by matching parts $b = 1/a_o$ and $E = -\hbar^2/2ma_o^2 = -13.6eV$. In spherical coordinates, this is aided by using $\nabla^2\psi = r^{-2}\partial/\partial r(r^2\partial\psi/\partial r)$. $V = -Ze^2/4\pi\epsilon_o r$, and $hc = 12.4keV\mathring{A}$. The proper coefficient A is found by normalizing the wavefunction and using the definite integral from 0 to ∞ of $r^2e^{-cr}dr = 2/c^3$. Already knowing the form of the solution is of course a big advantage.

Strangely, I had never been taught this in any classes. Dealing with complexity and generality sometimes pre-empts understanding things simply. Einstein advocated attempting a dual approach where any correct complex idea should also be explained simply (as to

[4] Actually, correct normalization has to include the overlap integral $\Delta = \int \psi_\ell\psi_r dV$ to give a coefficient of $1/\sqrt{2(1\mp\Delta)}$ [10]. That makes the splitting asymmetrical.

a 'barmaid' or to a 'grandmother' but now more appropriately "to a high school student"). We are used to not being able to describe an electron particle in the 1S ground state. But the further question is, "What is the electron **wave** doing in this ground state?"

Many texts on quantum mechanics include some explanation of the orbitals of the hydrogen atom. They are generally understandable until they discuss the radial portion of the wave-function, $R_{n\ell}(r)$ in $\psi_{n\ell m} = R_{n\ell}(r)Y_\ell^m(\theta,\phi) = (1/r)u_{n\ell}Y_\ell^m$ (where u is called the 'reduced radial function' and the Y's are spherical harmonics $-$ the vibrating modes of a spherical surface). Here, we are less concerned with the angular contribution and set $\ell = 0, m = 0$. The radial wave equation is often expressed in terms of 'Laguerre polynomials' but with a variety of differing conventions being used. Sometimes, authors avoid these polynomials and just use power series solutions or even hypergeometric functions. Students then often view even the simplest radial functions as mysterious because of uneven and poorly presented heuristics and lack of simplifying explanations. Chemistry texts and even physical chemistry books are even worse by freely using the names 'S-orbitals' or their 'σ-bonds without deriving or clearly explaining them.

If a text bothers to list Laguerre polynomials, they usually begin with: $L_0 = 1, L_1 = 1-\rho$ where $L_j = e^\rho(d/d\rho)^j(\rho^j e^{-\rho})$ [Rodrigues]. The 'generalized Laguerre polynomials' also connect to the radial locations of the angular functions [5] so that:

$$u_{n\ell} = N_{n\ell}\rho^{\ell+1}L_{n-\ell-1}^{2\ell+1}(\rho)e^{-\rho/2}, \qquad \rho = 2Zr/na_o. \tag{4}$$

Without that L_{n-1} subscript, one cannot connect to the form L_o for u_{10} where the first n value is 1 rather than 0 . Now we can see that the forms for ψ_1 and ψ_2 in (3) could include L_o and L_1. The 1S orbital wavefunction amplitude is an exponential decay away from the center of mass of the electron-proton system. The 'probability of finding an electron at a radial location r is given by $P = \psi^*\psi$'. The $V \propto -1/r$ Coulomb potential constricts the wavefunction towards the proton, but quantum mechanics also allows some exponential decaying probability of penetrating into the potential. In the ground state of hydrogen, the probability that the electron is inside the Bohr radius is only about 32% [6]. Ideally, one might ask the question, "what is the electron doing in the 1S orbital?" (or for that matter, in any orbital and in any chemical bond). There is no acceptable answer to this question. There is not even agreement that it is a legitimate question or even that an electron might exist prior to its being measured.

The old Bohr orbits could be pictured. After de Broglie, they represented standing waves that orbited in a plane and continually reinforced each other. The waves on the surface of a balloon can also be considered as reinforcing waves in both the theta and phi directions together. Can that be done for these new S-orbitals? No. They have a wide range of Fourier transform momenta representing a distribution of wavelengths superimposed to give a shape in space. In particular, the Fourier transform of $e^{-|x|}$ is a Lorentzian profile in 1D, and in 3D FT we have a Lorentzian squared:

$$e^{-r/r_o}/(Vol = 4\pi r_o^3/3) \quad \rightarrow \quad 6/(1+4\pi r_o^2 s^2)^2 \tag{5}$$

The decaying 'tent' profile of '1S' in space does imply something about implied momentum components via the uncertainty principle. And with the radial coupling to the spherical harmonics $Y_\ell^m(\theta,\phi)$, there must also be a distribution of radii and momenta for each of the separate spherical harmonics as well. The days of simple pictures are long gone.

Are these Laguerre polynomials necessary to understanding why the 1S orbital has exponential decaying amplitude? No. The Schrödinger equation represents conservation of energy in operator form: $p^2/2m+V=E$. But $p^2={p_x}^2+{p_y}^2+{p_z}^2$ and $V(r)=V(x,y,z)$, so much perspective can be gained from just considering the equation in one x-dimension. And a similiar exponential decay applies there as it does to the 3D central potential problem.

3. Analogies:

The simplest analogy is the one-dimensional particle in a box ($x=-a$ to $x=+a$). The lowest energy level is given by: $\psi=(1/2\sqrt{a})[(e^{ikx}+e^{-ikx})=2\cos(kx)]=\cos(kx)/\sqrt{a}$ where $k=2\pi/\lambda=\pi/2a$. The fixed f(x) shape is due to interference between left and right moving waves. The polar form is $\psi\equiv Re^{is\hbar}=(1/\sqrt{a})\cos(kx)\,e^{-iEt/\hbar}$. This can be generalized to 3D for a central cosine shaped wave peak in x,y,z. The 'left-and-right' moving interference in a 3D spherical cell might suggest 'in-and-out' moving radial waves.

The instructive case of a 'One-Dimensional Coulomb Problem' [7] or 'one-dimensional hydrogen atom' [8] central potential actually turns out to have some special complexities not found in the 3D case. It is in fact a controversial arena with offered claims and later refutations persisting at least to the 1980's. The potential $V(x)=e^2/4\pi\epsilon_o|x|$ has a singularity at $x=0$ which is the source of difficulty and allows no transmission through the origin between separate left and right wavefunction portions. These regular wavefunctions vanish at the origin unlike the 3D case which has a ground state peak there. The existing wavefunctions still use the associated Laguerre polynomials, L, and exponential decays to the left and right with decay constants $1/na_o$. The form of the functions are $\psi\propto xL(\rho)exp(-\rho/2)$ where the factor of x is needed to cancel out the -1/x potential. There are no eigenstates with definite parity. But, the problem does produce the usual **Balmer** series (lowest state is $n=2$) with the same energy spectrum as the 3D H-atom. So this case is a partial counter-example to 1D being simpler than 3D. Strangely, this problem also admits anomalous half-odd integral n states with even appearing wavefunctions more resembling those of the 3D hydrogen atom except for a narrow divot at x = 0.

The 3D 'spherical harmonic oscillator' ('SHO') and also the case for a spherical box potential provide relevant examples for contemplation. Note that a three dimensional spherical isotropic harmonic oscillator also uses Laguerre polynomials in their wave function solutions [9]. The ground state in this case is a centralized **Gaussian**, $\psi_o\sim\exp(-r^2/2)$

which is then a 'kin' to the atomic S-wave. How does this state have any kinetic energy? [5] Also, the FT of a Gaussian is also a Gaussian in 3D for the SHO, and of course that is also true in 1D. So we don't have a nice picture somewhat related to a closed Bohr orbital standing wavelength − but rather a distribution of momenta. Similarly, the classical 2D 'drum head' and 3D 'spherical resonance cavity' are characterized by Bessel functions J_o and j_o with 'Radial FTs ' which are also distributions. The ground state of a one-dimensional LHO uses the Hermite polynomial $H_o(x) = 1$ and is also a Gaussian. The spherical square well potential also has a spherical Bessel function solutions, e.g., $j_o = \sin(\rho)/\rho$ (like the 'sinc' function) where $\rho = \alpha r$, and $\alpha\hbar = \sqrt{2m(V-E)}$.

So, the potential well determines the location and momentum constraints on the ground state values. The electron wavefunction can penetrate the potential barrier as a decaying tail. The inverse square field is strong enough so that the ground S state only possesses this exponential decay character. In contrast, the spherical harmonic oscillator parabola potential is soft enough so that the ground state can develop more character and end up with a Gaussian bell-shaped profile. These both correspond to the first Laguerre polynomial, L_o (so there is no special mysterious tie-in).

For the commonplace LHO problem (linear harmonic oscillator with $V = kx^2/2$), the ground state Gaussian wavefunction is centrally located:

$$u_o(x) = A\exp(-\alpha^2x^2/2) = \frac{\alpha^{1/2}}{\pi^{1/4}}e^{-\alpha^2x^2/2}, \quad \alpha^4 = km/\hbar^2. \tag{6}$$

The expectation values for $< x >$ and $< p >$ are both zero (because they are odd functions of x). The expectation values $< x^2 > = 1/2\alpha^2$ and $< p^2 > = \hbar^2\alpha^2/2$. Since expectation values for Delta x and Delta p are given by variances, $\Delta x\Delta p = \sqrt{< x^2 >< p^2 >} = \hbar/2$, the tightest uncertainty. For the next state $u_1(x) \propto 2\alpha x e^{-\alpha^2x^2/2}$, $\Delta x\Delta p = 3\hbar/2$ [4].
Notice that the central portion of the LHO or SHO wavefunction is smooth (mid Gaussian) because matter wave forces vanish at zero radius. But, for the hydrogen atom with inverse square field, the potential and forces become infinite at zero radius. In this case the wavefunction is not smooth (it is a peaked exponential decay from center).

4. Discussion

A common curiosity about introductory derivations for the one-electron atom is being able to discuss and use a central potential from a nucleus to well defined electron locations. An electron as a particle cannot be localized to within about one Bohr radius, a_o, due to the uncertainty principle. But the electrostatic potential is given for a particle with definite precise radial location. The unlikely interpretation might be called "Whack-a-mole" (a board game in which a mole sticks its head out of a circle and then gets whacked with a

[5] KE could come from the usual formula $-\hbar^2\nabla^2\psi/2m$, but again Bohm would have a motionless electron with no KE. Although a minority view, the pilot-wave interpretation advocates are increasing in number.

hammer only to have another mole pop up from another hole, etc.). It is as if single electrons suddenly materialize in accordance with the Born probability and then vanish only to appear again at another location until all locations experience the materializations. A similar problem occurs in many other examples such as the derivation of the van der Waals interaction which uses potentials for two electrons in two atoms as if each atom possessed an instantaneous dipole moment for dipole-dipole interactions. An old belief was that the electrons zip around very quickly so that they can have instantaneous positions but still effectively cover a diffuse cloud. A modern belief is that quantum mechanics describes waves only, and quantum field theory describes fields and perturbations of fields only without actual existence of localized particles. An actual whole electron charge doesn't have to exist everywhere because the quantum-electron-field existing everywhere contains knowledge of the electron charge along with its other properties. Field interactions can use that knowledge in their processings.

Using specific radii makes sense if one treats space-time as possessing mathematical mesh 'cells' of values to be updated. The potential 'conditions' the space. In non-relativistic quantum mechanics (NR-QM), each <u>cell</u> has a specific location. For electrostatic fields, the entity to update iteratively is the EM potential such that the Laplacian of U is: $\nabla^2 U = -\rho/\epsilon_o$. In free space outside of charge sources, the Laplacian can be considered to represent the process of iterative averaging of the values $U(x, y, z, t)$ of a cell over the values in the nearest neighbors. Rather than solving the problem long range over space-time, the process is merely local updating by iterative averaging and continuing these averagings over cells until given boundary conditions (BC's) are satisfied. The boundary conditions propagate their values to the cell. The EM values of the cell are treated separately from an electron which might actually occupy the cell. The same applies to Newtonian gravitation, $\nabla^2\phi = 4\pi G\rho$ (in for example a neutron crystal interferometer experiment).

The physical interpretation of Poisson's equation with sources is numerically a little more difficult. The quantum mechanical problem for say a one-electron atom is still more difficult: $H_{rel}\psi_n = E_n u_n$ or $\nabla^2\psi = -2Z(r/a_o)\psi$. And, in this case, each cell possesses an electromagnetic potential value, U, and also a separate and possibly complex quantum mechanical amplitude value, $\psi(x, y, z)$.

No one really understands the particle property of 'charge;' its origin and characteristics lie beyond the standard model. There is an intuitive discrepancy between the particle picture (full charge instantaneously at each location along with a Born-Oppenheimer approximation) and Schrödinger's old idea of a diffuse cloud charge density with partial charges, $dQ = e\psi^*\psi \, dV$ [6]. The wave function is supposed to contain all knowledge, so extend that to knowledge of charge also. The wave function *IS* the particle and with the right

[6] For consistency, note that the potential energy of a 1S orbital for a nucleus of charge Q_N has:

$$\langle V\rangle = \langle\psi|V|\psi\rangle = \int \psi^*\psi\frac{Q_N Q_e}{4\pi\epsilon_o r}d(vol) = \int \frac{Q_N}{4\pi\epsilon_o r}(\psi^*\psi Q_e)d(vol) = \int \frac{Q_N}{4\pi\epsilon_o r}\frac{dQ_e}{d(vol)}d(vol) = \int \frac{Q_N}{4\pi\epsilon_o}\frac{dQ_e(r)}{r}.$$

Hamiltonian represents everything physical particles would do. The electron field in QFT is understood to contain knowledge of electron properties over all space-time. My perspective is to assume that space-time processes all these particle locations and potential interactions as a simulation of all interactions prior to 'final result.' A time-independent standing wave continually self-reinforcement aids the 'materialization' of active partial charge in electron clouds and overlapping electron clouds.[7] They acquire a more 'real' status than just ψ but less status than that of a discrete measurement. This charge excess behaves as a source of attraction and interacts with both positive nuclei. This behavior is similar to usual classical electrostatic attraction. So the quantum overlap integral has taken one intermediate step towards becoming classical. The reality of this overlap-excess is apparent independently of active observation. The molecules in a room would fly apart and explode without the reality of chemical bonding from quantum effects.

We said that the 1S single atom ground state amplitude has an oscillation in time like a tenting shape which points up and then points down and then up again. This is like the lowest mode of a drumhead which rounds up and then depresses down and then up again for the lowest sound wave. The molecular orbital (MO)-wavefunction also vibrates in time due to the energy of the system. So, an ${H_2}^+$ or H_2 molecule has a ψ that looks like a suspension bridge which faces up, then inverts itself down, and then up again with time.

How about hydrogen atom angular momentum orbitals with waves going both 'forward' and 'backwards?' Two of the lowest Legendre polynomials are $P_1 = \cos(\theta)$ and $P_2 = (3\cos^2(\theta) - 1)/2$. We could rewrite these as $P_1 = (e^{+i\theta} + e^{-i\theta})/2 = \cos(\theta)$ representing a superposition of a wave in the positive and negative theta directions. And $P_2 = (3\cos^2(\theta) - 1)/2 = (3/4)\cos(2\theta) + (1/4)$, where $\cos(2\theta) = [e^{+i2\theta} + e^{-i2\theta}]/2$. This again resembles a fixed shape due to interference between forward and backward moving waves where theta is some omega t: $\theta = \omega t$.

Note that physicists and chemists express some orbitals differently. The Legendre polynomial for $\ell = 1, m = 1$ is $P_1^1(\cos\theta) = (1-[\cos\theta]^2)^{1/2}$, but that is just $\sin\theta$. Then physicists write $u_{21\pm1} \propto \sin\theta e^{\pm i\phi}$; and chemists write $\psi_{2pz} \propto \cos\theta$ but also $\psi_{2px} \propto \sin\theta\cos\phi$ and $\psi_{2py} \propto \sin\theta\sin\phi$. Which is OK since $e^{\pm i\phi} = \cos\phi \pm i\sin\phi$. This allows chemists their p-"lobes" with one side having plus amplitude and the other having minus amplitudes for a labeled figure-8 picture. The usual "p-lobe" pictures are for amplitude squared − but does that really occur prior to interaction with another atom? When does the Born rule occur? If a $2p_x$ plus side amplitude lobe combines with a a 1S atom orbital, the electron density in that side is enhanced so that the effective size of the opposite unused p-lobe is

For the hydrogen atom with a nucleus of just one proton, this becomes $\langle V\rangle = -\hbar^2/{a_o}^2 m_e \simeq -27.2$ eV (one hartree). This charge density view is not very useful for the time dependent moving electron case, and there is no repeating reinforcement there. But it seems to be true here. Also note that if $\psi^*\psi$ suddenly ceased, you and all your surroundings would suddenly explode.

[7]How much reinforcement is needed? Perhaps there is some characteristic time constant τ for each system so that an adequate time can be expressed as a fraction of unity by $(2/\pi)\tan^{-1}(t/\tau)$.

diminished. The Born rule changes the density of the electron cloud.

In QM, it is permissible to linearly combine base states with coefficients which can be complex to obtain new candidate wavefunctions. The ground state of carbon with its four outer electrons in shell '2' can recombine its 2S and 2p orbitals as follows: $1s^2 2s^2 2p^2 \rightarrow 1s^2(2s^1 2p_x^1 2p_y^1 2p_z^1)$ [12]. And then these four outer electrons can then be added or subtracted together to give 'tetrahedral hybridization, sp^3 ' $-$ (e.g., a lobe $s + p_x + p_y + p_z$ in the $\hat{i} + \hat{j} + \hat{k}$ direction). These orbitals all had about the same energy, so promotion of one 2s electron is a minor change. Each of the equivalent sp^3 new orbitals has the same size, shape, and energy. Depending on chemical need and lowest energy, other hybrids could be formed. Chemical bonds do not have to be localized at the ends of lobes. For example benzine has strongly delocalized electrons in $\pi-$ bonds near all six of the 6C ring.

One implication of the 1D hydrogen atom to the 3D S wave is that one should not think of a particle or wave passing directly through the singularity at the proton nucleus. The expectation value of $< p^2 >$ for the 1S state is calculated to be $\hbar^2/a_0^2$ and $< x^2 >= 3a_o^2$. So, $\Delta x \Delta p = \sqrt{< x^2 >< p^2 >} = \sqrt{3}\,\hbar$. The expected kinetic energy is
$< KE >=< p^2 > /2m = +\hbar^2/2ma_o{}^2 \simeq +13.6eV$. But the expectation value of potential $< V >=< -e^2/4\pi\epsilon_o r >=< -\hbar^2/a_o mr >= -\hbar^2/a_o{}^2 m \simeq -27.4$ eV. So the net energy of the ground 1S of hydrogen is again $E \simeq -13.6eV$. This is just a special example of the virial theorem that $\langle T \rangle = -\langle V \rangle/2$ with $\langle (1/r) \rangle \propto 1/n^2$, or:

$$\langle \psi | T(p) | \psi \rangle = (\lambda/2) \langle \psi | V(r) | \psi \rangle \tag{7}$$

where the potential V is of degree $\lambda = 1$ here.
This is fairly straightforward. But it is difficult to discuss what the kinetic energy is like when atomic orbitals superimpose.

Measurements for long-distance entanglements are most easily understood by the Cramer 'backwards in time' transactional interpretation ('TI') of QM [13]. The discussion is for time-dependent Schrödinger's equation $-$ but what about the bound state time-independent Schrödinger equation? Could these transactions also occur in the short-distance entanglement of chemical bonding? Well, there would no longer be the usual 'sources and sinks', but there could be communication links between different 'space-time cells' (sub-quantum-mechanics). Certainly, QM for the more macro world of sources and sinks must derive from a sub-quantum world; and 'TI' could derive from a 'sub-TI' handshaking agreements across cells. We think of stationary-state orbitals and bonds in terms of back-and-forth motion of waves. If there were back-and-forth communication in time, it might be hard to tell the difference. Cramer theory 'derives' the Born rule $\psi^* \psi$ as a handshaking agreement between an offer wave ψ from a source and a verify wave moving backwards in time from a receiver sink to the source, ψ^*. Could it be that the Born rule derives in general from reinforcements that include backwards in time verify wave components?

I am tempted to define a new word, 'Qureal' or 'quantum real' to refers to a state of being part way between the classical world of observations and the quantum world of possibilities.[8] And the particular example is covalent bonding where the enhancement of overlap behaves as a Coulomb source of negative charge between nuclei. These time invariant standing waves represent a reality below the 'possibilist world' of TI by Ruth Kastner [14]. Although entanglement has been verified many times using the polarization of photons, it has not yet been verified for electrons (for example, electron spin). Most people believe in it, and testing may be done in the near future. TI can use psi-star for back in time verification for light because a photon is its own anti-particle. But electrons going back in time are positrons and move at sub-light speeds. TI needs to elaborate on its mechanisms for the case of massive particles (or matter waves).

References

[1] S. Nordholm, G. Bacskay, "The role of Quantum Dynamics in covalent bonding – A comparison of the Thomas-Fermi and Hückel Models," http://cdn.intechopen.com/pdfs/28320/InTech-The_role_of_quantum_dynamics_in_covalent_bonding_a_comparison_of_the_thomas_fermi_and_h_ckel_models.pdf

[2] K. Ruedenberg, Rev. Mod. Phys. 1962, 34, 326-352. Also see: Ruedenberg, K. and Schmidt, M.W. (2007), "Why does electron sharing lead to covalent bonding? A variational analysis," J. Comput. Chem., 28: 391-410. And see supporting comments by Frank Rioux.

[3] Richard Feynman, R. Leighton, M. Sands, The Feynman Lectures on Physics, III Quantum Mechanics, Chapter 10, Addison-Wesley, 1965.

[4] Robert B. Leighton, Principles of Modern Physics, McGraw-Hill, 1959.

[5] "Atomic Hydrogen and Hydrogenic Atoms," August 17, 2010, http://eg.bucknell.edu/ ∼ koutslts/Ph332/Text/Chap5_JQPMaster.pdf

[6] David McMahon, Quantum Mechanics demystified, McGraw-Hill, 2006.

[7] G. Abramovici, Y. Avishai, "The one-dimensional Coulomb Problem," arXiv:0905.3978 quant-ph 25 May 2009 or see J. Physics. A: Math. Theor. 2009. 42285302

[8] H.N. Nunez Yepez, et. al., "The one-dimensional hydrogen atom in momentum representation," Eur. J.Phys 8 (1987) 189-193 IOP-UK.

[9] John L. Powell, Bernd Crasemann, Quantum Mechanics, Addison-Wesley Publishing Co. 1961, 1965. Page 206+.

[10] Louis Harris, Arthur L. Loeb, Introduction to Wave Mechanics, McGraw-Hill, 1963.

[11] P.W. Atkins, Physical Chemistry, W.H. Freeman, 1978 p 461.

[12] Walter J. Moore, Physical Chemistry, Third Ed., Prentice-Hall, 1962.

[13] John G. Cramer, "The Transactional Interpretation of Quantum Mechanics," Rev. Mod. Physics, Vol 58, No. 3, July 1986, pp 647-687.

[14] Ruth E. Kastner, "The Possibilist Transactional Interpretation and Relativity," arXiv 1204.5227, 22 April 2012. (Intended mainly for measurements 'in which quanta are tied to sources and sinks.'). She has a new book on TI coming out in December 2012 from Cambridge Press.

[8] The term 'semi-classical is already taken (e.g., meaning WKB or limit as $\hbar \to 0$).

Reality of Schrodinger Ψ-Waves?

Dave Peterson, 9/8/18 – 9/14/18

For the last 80 years, physicists have pondered whether Schrodinger quantum waves might be "real;" and if so, then what kind of field are they made of ? (… for now just call it a "matter field"). Conventional positivistic Copenhagen dogma emphatically says "There is no quantum reality." But the community view has been gradually changing over time against Copenhagen. Now we have ongoing and strongly felt "interpretation wars" between "ontology versus epistemology" --meaning that some sort of quantum reality does exist versus the standard claim that quantum mechanics is merely a theory for calculating experimental outcomes (instrumentialism). A partial truth might lie in-between – a rarely considered "quantum omelet" with both egg and cheese mixed together.

This note discusses whether Schrodinger waves are even the right topic to consider. Spatial waves carrying momentum $p = h/\lambda$ are merely postulated in quantum mechanics [QM = non-relativistic quantum mechanics]. I call the joint starting foundation $E = hf$ and $p = h/\lambda$ as "Postulate Zero" for quantum mechanics [DP_2015]. The textbook postulates prefer to be in the form of operators: $p = -i\hbar\nabla$ and $E = i\hbar\, \partial / \partial t$, but these operators draw out momentum as a density of waves in space and energy as a density of waves in time (so postulate zero is there). But in relativistic RQM, **$p=h/\lambda$ is derived** from a more fundamental rest mass vibration $f_o = m_o c^2/h$. These considerations are discarded in standard QM thus making philosophical discussions somewhat "off-target." By itself, the $p=h/\lambda$ formula is still valid since it is a low speed case of the more encompassing special relativistic theory. Note that the momentum wave is not something possessed by the particle. A double-slit apparatus moving towards a particle "at rest" would see the same interference pattern as when it is the particle that moves [Shuler].

In relativistic QM, the nature and origin of wavelength λ **is due to loss of clock synchronization** seen by moving observers—a purely relativistic effect. In other words, it can be viewed as an artifact of Lorentz transformations between frames of reference. Although well known historically (e.g., de Broglie, 1924), this concept is almost never considered in interpretations of QM. Suppose that a "particle" is really a localized bundle of electron field quanta with some spatial extent and that all "parts" of this field are synchronized perfectly in phase with each other. The usual 1S-orbital of the hydrogen atom shares this trait; but its exponential decay profile may have a tighter degree of localization. The field extent can be considered as an array of very tiny clocks with their "second hands" moving incredibly quickly—like 10^{20} revolutions per second! An observer with relative motion, velocity V, sees these clocks as skewed [ref.Shu] from the "leading edge" to the "trailing edge" (say distance "X"). This means that a number of complete cycles difference could be seen over the extent. In relativity, relative motion shortens lengths, so let $\ell = X/\gamma$ where Lorentz-factor gamma ≥ 1. Then signals from the sides towards the "middle of the particle" will appear to propagate with speed c+V from one side and c-V from the other [ref.FOW]. The time difference between these two signals will be seen as $\Delta t = VX\gamma/c^2$, a term in a Lorentz transformation. It is a clock de-synchronization term that would always be zero for classical physics (with effectively infinite speed of light) but is non-zero for relativity. During this time difference, there can be several cycles of clock rotation difference between front and back – and this continuous phase difference becomes space wavelengths.

To understand this better, we first might be inclined (wrongly) to think simply of a moving electron as a "traveling vibrator" forming a spatial wave – but this is not a de Broglie wave! We would have $\lambda_{classical}$ = V/f_o which is very different from λ_{dB} =h/mV=h/p. To be relativistically correct, the "classical" wave would require phase velocity $\lambda f_o = v_\phi$ instead of V (and $v_\phi = c^2/V$ in special relativity -- as discussed below).

As an **example**, consider a 1 keV electron (here meaning KE = 1000 eV/c^2) with momentum p = $(2mE)^{1/2}$ = 32 keV/c traveling with a de Broglie wavelength λ =h/p =hc/pc = 0.387 angstroms [where h = 4.135×10^{-15} eV·sec. or hc = 12.4 keV Å]. Electron mass-energy is 511 keV/c^2 giving a rest frequency of $f_o = m_oc^2/h = 1.26\times10^{20}$ Hz (126 exa-hertz!). And the speed of the electron is about V=0.063 c. Then the wrong-formula λ = V/f_o = 1.5×10^{-13} m = 150 fm which is 258=c^2/V^2 times smaller than the de Broglie wavelength from special relativistic Lorentz transformations of a base frequency.

Timeline Background:

1. Planck's Constant, **h**: is a term in Wien's Law for black-bodies [1896] tested by Paschen [1897], and then as discrete Planck bundles E=hf [1900] and Einstein [1905 – photoelectric effect].
2. Light has momentum: p = h/λ = $\hbar$k = hf/c=E/c by Stark [1909] and Einstein [1916]. (k= 2π/λ and $\hbar \equiv h/2\pi$).
3. **de Broglie** electron wave momentum p=h/λ came from special relativity [1923-1924] and explained the Bohr atom as having integral number of wavelengths around circular orbits, $n\lambda = 2\pi r_n$.
4. Birth of non-relativistic Wave and Matrix Quantum Mechanics, 1926
5. Alternative QM: Bohmian "non-local hidden variable" theory [by de Broglie, 1927].
6. Formulas from de Broglie's Nobel Prize Lecture, 1929: the electron has a **rest mass vibration** $E_o=hf_o=m_oc^2$. Then he applied special relativity equations: f = γf_o=E/h, p = $\gamma m_o V$, $V\cdot v_{phase} = c^2$. So: wave momentum p = $\gamma m_o V = EV/c^2$ = hf/v_{phase} = h/λ. [See Appendix for proper Lorentz transformations].

7. Electron matter wave self interference: first seen in 1954 using an "electron biprism" which was just a thin charged wire crossing an electron beam (gold coated 3 micron spider web strand— [ref. AA]). Then an e-beam double slit experiment showed interference in 1961 and one-at-a-time single electron interference in 1989.

Discussion:

A preferred definition of "real" stresses things that are not so dependent on relative motion between object and observer. Then, the most likely fundamental quantum wave reality is that all small massive particles have a very rapid time varying scalar phase vibration representing energy $E_o=hf_o = m_oc^2$ – and that is indeed a relativistic invariant. This is also true for even large composite particles (like C_{60} and larger molecules), but a detailed explanation has not yet been made clear.

A big problem is that the Schrodinger wave p=h/λ being non-relativistic ignores any rest mass vibration, has a strange relation between frequency and wave-number

given by v_ϕ =phase speed = $f\lambda$ = velocity/2 (phase-speed lags real speed, and its kinetic energy $KE=p^2/2m$). It inherits $p=h/\lambda$ but then throws away its relativity parent. In special relativity, total energy **$E^2= (m_oc^2)^2+(pc)^2$.** *[see triangle below]*
If it is associated with waves with wave-number $k = 2\pi/\lambda$, then the energy formula can be re-written as: $(\hbar\omega)^2=(\hbar\omega_o)^2 + (c\hbar k)^2$. Then a standard formula for "group velocity" $V_{particle} = \partial\omega/\partial k = kc^2/\omega = c^2/v_\phi$ with "phase speed" $v_\phi= \omega/k = f\lambda$. So the product is $V \cdot v_\phi =c^2$. [We won't discuss here the strangeness that one has to square what are called "probability amplitudes" to get to actual measured results – the mysterious "Born Rule" that no one has yet derived clearly]. In the Pythagorean triangle above, as well as a created wavelength, the base frequency is also altered to a higher frequency seen by an observer.

We could say that we are addressing the topic of quantum reality with respect to the wrong ball-game. Relativity is a fundamental requirement. Ordinary quantum theory does address light waves; but their speed, c, is obviously relativistic—so we have to deliberately restrict discussion about them in conventional QM (and for light, both phase and group particle speed = λf = c). The Schrodinger space phase wave $p=h/\lambda$ depends on relative motion and so has a lesser degree of reality than something that is invariant. But, as humans, we often choose to say that what counts is that it is <u>real to us</u> in our "lab frame." (!) And QM focuses on what "the observer" sees.

It is a misfortune of history that de Broglie's math is rarely taught to students – largely because conventional wave mechanics quickly followed in time, was incredibly successful, and then effectively dominated over de Broglie's views. Relativity could be set aside for special use in relativistic QM and quantum field theory (QFT). Also, de Broglie math was first printed in French in 1925 without heuristic polish and was not translated into English for 80 years.

Here is the basic picture (imagine right triangles with better drawn hypotenuse – connect the $_0$'s with line segments):

<u>Energy Momentum triangle</u>

```
  hypotenuse                  0
   E = mc² =γmₒc²             |
  angle θ=arctan(γβ)          | pc
/0____________________________0
      Eₒ=mₒc²       =hfₒ
```

<u>Frequency right triangle</u>

```
                                  0
hypotenuse = f =γ fₒ              | side
  / angle θ                       | =c/λ= pc/h
/0________________________________0
Base= fₒ = "rest frequency"
```

These sketches show that as relative velocity β =V/c increases, angle θ increases and transforms the rest frame base to now include momentum and wave-number seen by the moving frame. The relativistic factor gamma will often be nearly $\gamma \simeq 1$. So, the production of de Broglie wavelength is cause by an upward rotation of the base into newly created space waves. A momentum 4-vector and a wave-number 4-vector transform the same way and can be considered as proportional using Planck's constant, h or $\hbar$ =h/2π . The Pythagorean theorem for a right triangle says that new total (moving) frequency $f^2 = f_o^2+(c/\lambda)^2$.

In relativity, all inertial frames are equally valid. Apart from Bragg electron diffraction from crystals, a standard test for the presence of spatial waves is the resulting interference pattern from electrons passing **through a double slit** (a focus of the recently selected book: Through Two Doors at Once). But the **same** interference pattern will be seen if the electron is considered at rest and the double slit apparatus with screen approaches the electron at some speed V (**!**) [Schu]. Now an electron at rest has an infinite wavelength (λ = h/0 = ∞) —so how can it interfere? The answer again is that only relative motion matters. de Broglie Wavelength is not a distance in the sense of a ruler in space! From crystal or slit diffraction, we cannot deduce backwards that a prior phase speed or phase wavelength has fundamental existence.

Matter Waves such as electrons can split into two paths and interfere with themselves [AA]. We cannot say that an electron “particle” travels both paths. It must be that an electron is a localized matter-field that delocalizes at it travels and finally gets localized again at the end [Wheeler’s “Great Smoky Dragon” with well defined head and tail but vague “smoky” middle]. The wave can take many paths through space-time. The end result is again a localization associated either with an (incomprehensible) wave function “collapse” or an early determined Bohmian spatial preference from initial conditions. Neither is very satisfactory. Bohm mechanics needs no collapse but depends on a “quantum potential” and an environmental wave function producing non-local effects. The establishment of such a wavefunction is mysterious. It is as if the electron knows where it is going to go because “it has already been there.”

Conclusion:

So, are Schrodinger’s ψ waves real? They exist only as a relative de-synchronization of the primary at-rest matter wave vibrations and so themselves lack fundamental reality. They act as if they are somehow **real to us** in our Lab frame and participate in basic 4-vector transformations (E,**p**) or (ω,**k**). The observer enters in the sense of establishing a relative speed between particle and detector/observer, and an observer at a different relative speed would see a different wavelength (it is relational). One cannot make a Lorentz transformation of the wave-length because it is already the result of Lorentz transformations. Those equations act to effectively rotate a base frequency up into an effective wavelength.

AfterThoughts:

The spatial extension of the electron field for a single particle is very strange and interesting. For low momentum, the wavelength can be very long. In quantum field theory, the quantum electron-field permeates all of space-time; and a “particle” is mainly a localization of the excitation of the field (a quanta). A particle is not a little spinning ball. A current “meme” aiding this understanding is, “There are no particles, there are only fields.” Strong localization occurs at particle creation and particle annihilation at the end – but with a very diffuse in-between. Localization also occurs during interactions such as the vertex of a Feynman diagram. Probing an electron at high energy localizes its field, and then we can say that the electron seems “point-like.” Another thing that forces localization is the exclusion principle. It says that if an electron in a region possesses certain quantum numbers, then any other electron with the same quantum numbers is not allowed to intrude or overlap the same region of space-time. This is especially interesting in white dwarf stars. Explanation from first principles is still unsatisfactory but shows that the concept of electron fermion spin and the exclusion principle are consistent (the “spin-statistics” theorem).

$p=h/\lambda$ comes from relativity. But why and how does relativity work? Why should all observers see a constant speed of light regardless of motion? For us, it is just a given: physical reality has to be Lorentz invariant. The "why" of relativity carries over to the "how" are conservation laws enforced? Energy, momentum, angular momentum and quantum numbers have work out upon detection/annihilation, and the mechanisms are unclear (more givens). How does entanglement occur (forces conservation laws) --and distance seems not to matter. Exactly what is a "matter wave." How do multi-particle wave functions operate in configuration space? In our current understanding, after a long distance, remaining very very tiny ψ –amplitudes don't matter – somehow they still work. We are still a long way from understanding the foundations of physics. And we do live in a "preposterous universe."

References.

[Schu] See: Robert L. Shuler [NASA], "Common Pedagogical Issues with de Broglie Waves: Moving Double Slits, Composite Mass, and Clock Synchronization." 2015
https://www.hindawi.com/journals/physri/2015/895134/
[DP_2015] Learning Quantum Mechanics and Relativity, Dave Peterson,
http://sackett.net/LearningQuantumMechanics.pdf
[AA] Anil Ananthaswamy, Through Two Doors at Once, the elegant experiment that captures the enigma of our quantum reality, Dutton, 2018.
[FOW] Michael Fowler, "Special Relativity: Synchronizing Clocks,"
http://galileoandeinstein.physics.virginia.edu/lectures/synchronizing.html

Appendix notes:

1). Another approach to deriving $p=h/\lambda$ is using the two standard space-time Lorentz transformations: **1.** $\mathbf{t'=\gamma(t-Vx/c^2)}$ **and 2.** $\mathbf{x' = \gamma\ (x-Vt)}$

where frame L'(x',t') has relative velocity V with respect to a particle at rest in frame L(x,t). The particle (localized electron matter field) has intrinsic vibration $E_o=hf_o=m_oc^2$ in frame L being observed by frame L'. [The term $\gamma Vx/c^2$ tells how clocks are no longer synchronized with each other (so different parts of its extended vibration have different phases.]

Pick initial time t=0 and let wave period $\Delta t' = 1/f$ where $f = \gamma f_o = \gamma m_o c^2$.
Solve equation 1 for x: $\Delta x = \Delta t' c^2/\gamma V = Vhc^2/\gamma hf = hc^2/V\gamma m_o c^2 = h/\gamma p$ where $p = \gamma m_o V$.
And then use transformation 2. for $\Delta x' = \gamma\Delta x = \gamma h/\gamma p =$ $\mathbf{h/p = \lambda}$!
[the view from frame L' about a particle lin L]. Lambda is the distance between phase peaks because delta t' was chosen to be one over frequency.

An electron is waves of what?

Dave Peterson, 10/6/17 -11/1/19 (revised, preliminary)

What is an electron quantum mechanical psi-function made of ? [Jones].
What is a generic answer for what is vibrating common to all cases of $E = \hbar\omega$?
What is the fundamental reality or substance of so-called "matter waves" ?

Scope:

The class of interpretations of QM considered here are "realistic" and apply to each measured event rather than being "epistemological" where psi only represents the knowledge of an observer over an ensemble of data events. Examples of realism include the "de Broglie-Bohm" interpretation, the "de Broglie double solution," and the "transactional" back-and-forth-in-time interpretation. We don't really have to say what a quantum wave is – we are already familiar with its behaviors (although some of our beliefs might turn out to be wrong). It can be called a psi-wave (the ψ-wave of a ψ-field) lacking any substantial properties that exist in the classical world.

The de Broglie relations for matter waves represent information about matter. de Broglie relations are basic in two different arenas: Relativistic quantum mechanics (RQM) where $E = \hbar\omega = \gamma\, m_o c^2$ refers to an intrinsic frequency numerically representing total mass, and the momentum $p = \hbar k$ is due to relative motion clock-de-synchronization of a standing wave source ω_o-phase (an artifact of the Lorentz transformation). This is contrasted with non-relativistic QM which excludes rest mass-energy, has E = KE + PE, and where the formula $p= mv =h/\lambda$ is just an empirical fact and postulate.

"Matter-waves" or "energy-waves" or "information waves" or something else ("ψ-waves"): Matter is concentrated energy mostly localized spatially within some confined volume. That describes quark/gluon-composite protons and neutrons with $m_n = E/c^2$ where quark masses only contribute a tiny portion to the total mass. Although photons are also called "particles," they are massless; so we might wish to call their waves "energy-waves" instead. But, as discussed below, ψ -waves or de Broglie waves in general are clearly not energy waves, they are more like "waves of information" that can be decoded to give energy or momentum. But then we want to think that there must be some "substance" or field that carries that information. We can give it the name "ψ-field" with perhaps detailed physics to be defined in the future.

It might be intuitively tempting to say that the de Broglie wave is electromagnetic for an electron and for massless photons. Or perhaps an electron's mass-producing give-and-take of "hypercharge" in the "Higgs-field" makes an electron wave. Note that the Higgs field interactions leave charge and spin intact, so perhaps they don't "vibrate." Each unique type of quantum field can produce an elementary particle quanta with the attributes specified by that field. A quantum phonon wave is something quite different, mechanical vibrations of nuclei on a lattice. And then we have large composite neutral macro-molecules also demonstrating de Broglie waves with wavelength $\lambda =h/Mv$ where M is the sum of all the masses of its entities. These waves seem to be only energy/momentum information waves can that assist in guiding the particle.

This over-riding commonality of de Broglie relations for all of these cases is profound and implies a higher-level quantum-energy principle of the Vacuum and an expanded definition of "quantum energy" or "pre-energy." Whatever it is that forms energy, it is constrained and governed by an “energy supervisor” that controls the packaging and shipping of quanta. So, a quantum wave may be composed of multiple superimposed things together such as a pseudo-electromagnetic field, **A**, combined with hypercharge fluctuation, Y, but always also with a common “pre-energy-information wave.”

Massive quantum particles can have many attributes, the most important of which are the mechanical matter attributes of observable energy and momentum, E and p, encoded in what has been most commonly referred to as their quantum mechanical “matter waves.” Irrespective of amplitude, energy corresponds to the number of wave vibrations in time, and momentum is the density of wavelengths in space (the gradient of the phase, e.g., $\varphi = \mathbf{k}\cdot\mathbf{r} - \omega t$). For spherical or dipole waves, long distances can lead to extremely faint amplitudes. But they don’t seem to be lost in background noise and act like a “dedicated” wave for a particle.

The de Broglie relations of 1924, $E = \hbar\omega$ and $p = h/\lambda = \hbar k$, originally derived from relativity theory where E was intended to be the total mass/energy of an object. So, for 4-vectors (E,**p**) and (ω, **k**) we write $p^\mu = \hbar k^\mu$. Quantum energy is equivalent to the fundamental vibration of a mass in natural units – and we factor in a value for $\hbar$ that translates this to our “people sized” units like SI *(~MKSA).* In quantum mechanics with wave-functions, psi, we talk about an operator $\hat{E}$ on $\psi = i\hbar\, \partial \psi / \partial t$ and an operator $\hat{p}$ of $\psi = -i\hbar \nabla \psi$ that reveals the wave densities in time and in space; the operators are the space-time decoders of information in a wave.

In non-relativistic QM, energy E = KE + PE uses just kinetic energy, $KE = p^2/2m$, and totally ignores an underlying mass-energy. Foundationally, rest mass and its frequency, m_o and ω_o, are invariants of Lorentz transformations. Kinetic energy of motion $= (\gamma - 1)m_oc^2 = E - E_o$ for $v \ll c$ (so $E = E_o + KE$, and $\omega = \omega_o + \omega_{KE}$). It is interesting that quantum energy vibrations for rest mass versus KE can be decomposed and added together this way.

Establishing the classical theory of energy conservation was a long and difficult process. Even the idea of kinetic energy was absent in Newton’s 1687 Principia. But then we’ve added and interconverted electrical and electromagnetic energy, electromagnetic potential energy, gravitational potential, chemical, nuclear, thermal, sound energy, and wave energy. It wasn’t until 1850 that William Rankine first used the general phrase “the law of the conservation of energy.”

In the quantum world (QM), the Schrodinger Hamiltonian H = KE +PE ensures that a wave function solution of Hamiltonian operator $H\psi = E\psi$ will be consistent with this energy conservation. In quantum field theory, conservation of energy and momentum is dictated and imposed by stated delta functions such as $\delta^4(p_1' + p_2' - p_1 - p_2)$.

Particular Case Examples:

Single massless photons obey $p^\mu = \hbar k^\mu$ but are able to pass through, refract in, and reflect from complex glass arrays on optical benches in a way very much like a large electromagnetic wave interacting with all the electrons in the glass. We are thus tempted to say that single photon waves at least in part are electromagnetic. Beam splitters can split single-photon waves into parts that can interfere with each other later on. The final detection is again one single photon (and we can now detect and also emit single photons). Coherent superpositions of a large number of photons become more light-like with lower quantum fluctuations.

It is sometimes convenient and even effective to model a single photon as an electromagnetic wave such as the "Riemann-Silberstein" vector $\psi(x,t) \propto E+iB$, but there is no rigorous quantum mechanical justification to do so. Quantum and Classical are intrinsically different arenas, and there is no such thing as a position operator for a photon. Generically, it is then wrong to say that λ is only electromagnetic. It is at least also something else, a "matter wave" or "energy wave" unique to the quantum world. It is never exactly like anything with which we are familiar in our classical world, and it obeys different quantum rules *such as "wave-particle duality".* It can be said that "all quantum states comprise two physical components: one is the source of the energy (radiation, lattice vibrations, particles) and the other is the energy state" [Street]. We can call it a "ψ-wave" but it is not $\psi(x,t)$ with any definite position.

Electrons diffract from metal crystals as if they also might have electromagnetic waves somehow mapped from classical to quantum. They have an electric field from their charge, but diffraction comes from their quantum waves. Unlike the classical case, a fundamental quantum-vibration of an electron cannot radiate energy but does propagate ephemeral quantum waves .

We also have neutral neutrons diffracting through silicon crystal interferometers – certainly not from electric interactions (but their spin magnetic moments can interact with silicon nuclear spins).

An ultra-cold Bose-Einstein condensate (BEC) forms when the thermal de Broglie wavelength λ is near the interatomic separation so that neutral atom wavepackets collectively overlap their non-electromagnetic matter waves { -- the half-integral spin of alkaline nuclei and the half-integral spin of electron-shell can form an integral spin boson atom}.

Elelmentary particle muons and quarks (μ,u,d,s,c,b) also obey de-Broglie equations and can form temporary joint *"integral-wavelength"* composite-particle hydrogen-like orbitals with their antiparticles (ignoring spin, and m, and the 1S states). Examples are: e^+e^- positronium with $E_n \propto 1/n^2$, $\mu^-\mu^+$ dimuonium, the s strange quark combined with s-bar to make a ϕ-meson, c with c-bar charmonium or J/ψ like mesons, and b with b-bar "bottomonium" or "upsilon" mesons. The case for weak particles like neutrinos and W's and Z's is less clear, but they obey quantum field theory (QFT) which includes QM (mostly).

And finally, we have very eye-opening experimental examples of huge macro-molecules such as C^{60} "buckyballs" (1999) still obeying $\lambda = h/Mv$ where M is nearly the sum of all their atomic masses. The macromolecules showing laboratory interference today are much more massive than sixty carbons (e.g., **25,000 amu !** , [Arndt, 2019]). Despite tiny wavelengths much smaller than the molecules, the matter wave can pass through two slits much more widely spaced than molecular size. This is a challenge to realists.

Now, for electrons, we might imagine that an elaborated ψ wave contains enough information about electron-particles that an electron-wave passing through two slits might re-construct a physical electron at detection where "intensity of presence" $\psi^*\psi$ rules. It would be ridiculous to talk about end-point reconstruction for ultra-complex macro-molecules. A simple de Broglie wave passes through both slits, but the big particle must only pass through one of them. There is no re-assembly, the continuously moving particle hits the detector still as a particle "guided by the wave."

Matter waves can scatter off of other matter; and other interactions such as charge-charge are covered by the Hamiltonian energy conservation requirements. It could be that quantum waves carry more attributes than just energy and momentum (such as multiplication by a spin wavefunction), but E and p are the two covered by de Broglie rules. We now know that quantum waves for electrons and photons also can possess orbital angular momentum, ℓ, as well as spin angular momentum, s.

The de Broglie wavelength represents a center of mass degree of freedom ignoring all internal structure. For a classical composite system, total mass is obviously the sum of the individual masses: $M = \Sigma\, m_n$. But in de Broglie quantum mechanics we are considering the idea that composite particle frequencies result in a summing of their frequencies: $\omega_{total} = \Sigma\, \omega_n$. Here the idea of additive energy and additive mass with additive frequencies is somewhat strange, and "it appears that this remains an unsolved problem" [Shuler]. One author speculates that nonlinear interference effects may be needed.

For perspective, consider the case of positronium (which is similar to the hydrogen atom problem). A wavefunction for the combined electron (e) and positron (p) is the composite state: $\psi(r_e,r_p) = \psi_e(r_e)\psi_p(r_p) \rightarrow \psi_{rel}(r_e-r_p)\Phi_{cm}(R_{cm})$ for relative motion and center of mass motion. A product wavefunction is separable and allows separation of variables. Now, for simple de Broglie waves of the form $\varphi = \exp[(i/\hbar)(px-Et)]$, a product wavefunction $\Phi = \varphi_1\varphi_2 \rightarrow \exp[(i/\hbar)(\{p_1+p_2\}x - \{E_1+E_2\}t]$. If energy is the total mass/energy, then $E_1+E_2 = \omega_1 + \omega_2 = M_1+M_2$, the sum of component masses. But, admittedly, this is just a loose hand-waving argument that doesn't really come close to solving the problem.

So, it seems that the most fundamental reality is that all particles even including macromolecules with summed mass-energy ($E = \gamma mc^2 < 20$ TeV *current upper bound*) possess some sort of **fundamental vibrations**, $E = h\nu = \hbar\omega$ (or $E_o = h\nu_o = \hbar\omega_o$ for particles "at rest"). A logical upper limit is to this must be < $M_{planck} \approx 1.22\times10^{19}$ GeV/c^2 = 21.7 μg (and 20 TeV is 2×10^{13} eV).

Discussions:

Names, Properties and Use of Matter Waves:

The philosopher Ludwig Wittgenstein said, "The limits of my language are the limits of my world." Historically, and pragmatically, it was compulsory to only use classical words and a classical measuring apparatus in discussions on quantum mechanics. But the quantum world strongly differs from the classical world and demands its own terms. One cannot even begin to discuss any possible sub-quantum "reality" without using new words and making some attempt to define them adequately. Most importantly, the horribly ambiguous term "real" has to be broadened from classical physics to a separate "sub-quantum-real" at the level of psi, ψ rather than $|\psi|^2$. Most discussions of and articles on sub-quantum mechanics involve people talking past each other because they are unable to convey what they mean when they use inappropriate words. *{ My writing has always used the prefixes "pre-", "qu-" or "psi-" for: "pre-real", "qu-real,"* *"psi-real," "qu-wave," "qu-spin", or "psi-energy"}.* The physics philosopher Ruth Kastner uses terms like "quantumland" -- a large domain of "pre-spacetime" (pre-spatial-temporal or "PST"), sub-empirical possibilities for quantum states and their interactions. She refers to the "possibilist" world of ψ – a term essentially validated by Fermi's Golden rule where ψ knows about its end results being affected by a "density of final states" that is also used in the Purcell effect (validated about 1990 using a semiconductor micro-cavity). An atom in an excited state "cases-out" its entire environment before any actual emission takes place. It explores all possibilities for every final single event. In addition, in the relativistic realm, particles such as photons or electrons can come into existence – emerging from possibilities.

The orthodox meaning of ψ in Copenhagen quantum mechanics is "waves of probability amplitude;" an interpretation that is familiar but also numbingly opaque. The name is certainly appropriate "for all practical purposes" *[Bell's acronym is "FAPP"].* But that need not be the whole story. Matter waves could be "quantum-real" prior to measurement but eventually couple to a separate last stage action "Principle of random selection" or "stochastic choice" of ψ*ψ "intensity presence" -- a two-step process resulting in every particular "collapse" event. John Cramer's two step process is initial "offer waves" encountering possible receiver candidates which then broadcast quantum waves backwards in time to a source resulting in a "transaction." This has the virtue of explaining the mysterious Born Rule, ψ*ψ. Alternatively, Bohm might have been right with his "non-local hidden variables" that avoid the concept of collapse altogether. The standard view of psi is very unsatisfying – but few try to go deeper. No one is comfortable with "collapse."

Rather than just being postulated, the de Broglie matter-momentum wavelength λ =h/p derived from a Lorentz transformation of this rest-mass frequency due to relative motion, v, for the wave 4-vector $k^\mu = (\omega, \mathbf{k})$ *[Peterson]* . This matter-momentum wavelength λ =h/p might then be considered to have a lower (non-invariant) reality because it depends on the relative velocity, V, between source and observer. In 1924, Louis de Broglie said that matter wavelength λ represents a loss of clock synchronization seen by moving observers —a purely relativistic effect even at low speeds. That is, the Lorentz transformation of time has the term $\gamma Vx/c^2$ that tells how two clocks in different frames of reference differ in synchronization over distance. In other words, p=h/λ can be viewed as an artifact of Lorentz transformations between frames of reference and can be considered "real" only in a relevant frame of reference. The greater or invariant reality is the particle rest mass and rest frequency.

Quantum mechanics is presently a **theory of measurement** that has typically avoided discussion of any causal reality. The Schrodinger psi ψ(x,t) is just a solution of Hψ = (KE+V)ψ. What sort of things can go into a wavefunction: There are of course terms with scalar energy and momentum perhaps in terms of ω's and k's, amplitude fall-offs with distances and angles, angular momenta (maybe referring to quantum numbers like s, ℓ, j,n, m…things that might be conserved). Pauli matrices may be included for fermion spin, and there may be a vector term for photon polarizations. Mostly, these things do not refer explicitly to electromagnetism. Interactions of particles and fields are expressed via terms in the Hamiltonian rather than the wavefunction.

There is a broader version of the Schrodinger equation called the Pauli equation, and it can also include the electro-magnetic vector potential, A, spin in magnetic fields,

σ·B, and electric potentials, ϕ. But again, those are in the 1927 Hamiltonian rather than psi.
It is expressed as [Gurtler] :

$$H|\psi\rangle = \{(p-qA)^2/2m + q\phi - q\hbar\mu_n\boldsymbol{\sigma}\cdot\mathbf{B}/2m\}|\psi\rangle = i\hbar\,\partial\,|\psi\rangle/\partial\,t = E|\psi\rangle\ ,$$

where "sigma" means Pauli vector matrices, p is the operator ($-i\nabla/\hbar$), and $v \ll c$ spinors could be introduced as $\psi \rightarrow (\psi_+, \psi_-)$ if desired. For neutrons, the dipole moment $\mu_n\mu_N$ might only interact with a magnetic field, B (or "magnetic Bragg scattering" from a crystal lattice).

There have been and still are many confusion factors that muddy the waters of understanding quantum mechanics below the level of measurement. We don't yet have the right view – perhaps because a right view is unbelievably wild or too unbelievably obvious. And quantum field theory [QFT] doesn't add much clarity to quantum mechanics [QM] because these subjects differ in math and interpretations. In standard (non-relativistic) QM, observables are called "operators;" and Coulomb fields and measuring devices are classical. In quantum field theory (QFT), fields are basic and ψ's are operators for creation or annihilation of field quanta in various normal modes -- and the fields are chaotic. A quantum field is an entity existing at every point in space which regulates the creation and annihilation of particles. QFT identifies a wave with the superposition of an indefinite number of particles, and particle numbers are elementary excitations of their underlying quantum matter field.

And considering the photon again, the relations $E=\hbar\omega$ and $p = h/\lambda$ were found first for light quanta and then later for matter. Conventional electromagnetic waves do have energy due to electric field E*E's and magnetic field B^2's. And they can carry relativistic momentum and angular momentum. But there is no consistent way to add up the little bits of energy in each wave crest to obtain a full quanta of mass and momentum – a photon just isn't classical.

And for weak interactions, it isn't clear that the arena labeled as "electroweak" is a separate realm by itself. Kaku *(QFT p. 380)* says that the theory of leptons given by the Weinberg-Salam model is actually flawed by the presence of anomalies, and the true model requires quarks to cancel the anomalies. Anomalies can destroy renormalization. The photon apparently is a mix of W and B massless fields after Higgs-breaking. In that sense, the photon is electroweak. However, the "weak field" is usually thought of as just the massive vector bosons W by themselves.

Each type of quantum field permeates all of space-time for an overlapping set that Frank Wilczek calls "the **Grid**." The names of the various quantum fields separately include: (e, μ, τ) charged lepton fields, 3 types of neutral neutrino fields, 6 quark fields (u,d,s,c,b,t), EM photon field, electro-weak massive boson fields (Z^o, $W^\pm$ -- after electroweak symmetry breaking! {*EWSB}*), the Higgs field and 8 gluon boson fields. But counting them is hard because many of them exist in multiplets: the Higgs doublet, left-handed lepton doublets, right-handed electron singlets, both left and right handed quarks of various flavors and quark doublets, and gluon/color triplets. The Higgs field does not interact with gluons so that they have no mass.

What the “matter wave” might be depends strongly on a chosen interpretation of quantum mechanics. Having waves in configuration space, entanglement, superposition, complex numbers, and collapse weigh heavily against any conventional reality to the wave function and encourage epistemological interpretations (knowledge of ensemble behavior to an observer).

Elementary Particle Mass and Vibrations due to interactions with the vacuum expectation value (vev) of the all-pervading Higgs field:

Small elementary fermion particles are quanta of their separate quantum fields, and their masses theoretically come from exchanges of hypercharge in the universally permeating Higgs field *[see Discussion below].* But, overall, the interactions with the Higgs field only account for less than two-percent of the mass in the universe – much of the rest is mass-energy $M=E/c^2$. The frequency of this interchanging (called “zig-zags”) is the fundamental frequency *[see references at end].* Penrose addresses this as the Dirac equation coupling “two 2-spinors, each acting as a kind of source for the other” (~ “zitterbewegung”).

But, for all particle cases and masses, de Broglie waves are common to all regardless of origin; hence these waves are generic, and k^μ is often called just a “matter wave.” Its expanded interpretation is still a common topic of unresolved debate. Perhaps the hypercharge exchange with weak-isospin (ΔY_w vs. ΔT_3) transcends the fermion field names of the exchanges. But large composite “confined energy” particles need a way to sum all these (Y_w,T_3) frequencies up to the larger total matter wave frequency and with boson forces also contributing. A present conundrum is how to transition from Higgs field “zig-zags” to elementary particle quantum field mass/frequencies upwards to hadrons (localized confined energy) and then to large molecules. And then we need to be able to transition the concept of mass up to the cases of classical usage like bricks and planets and galaxies.

When we study electron mass from Higgs interactions, we learn that the electron is composite in “zig-zag’s” and that the rate of zig-zag goes with the mass of the particle (how well it couples to the Higgs field). The zig-zag rate for the electron is similar to $f = mc^2/h$ which is much smaller than that for the heavier t-quark (at 173 GeV). It **might** be that the “vibration” of the electron at rest is really “zig-zags” (e.g., [Tanedo],[Penrose], [Strassler]). And Zitterbewegung may be appropriate and real and relate to the mass of fermions.

Penrose adds that zigs correspond to the top “2-spinor” of the Dirac 4-spinor with helicity ½ $(1-\gamma^5)$ for <u>a left-handed</u> wavefunction, ψ_L and the zags have helicity ½ $(1+\gamma^5)$ for a right handed wavefunction, ψ_R. Only the zigs interact with weak particles W^+, W^-, Z^o and not the zags. Only the zigs go with the decay of the neutron fermion. For the positron, the reverse is true—only the zags. The neutrino also is a left-handed particle with only a zig. Now we know that the neutrino has a wavelength $\lambda = hc/E$ for energy with only a zig, so its oscillation is not due to zig-zags (some physics beyond that).

In terms of the third component of weak isospin and weak hypercharge, this is what takes place when an electron bumps into “the Higgs”:

e_L (T_3= $^1/_2$,Y=-1) interacts with the Higgs field condensate which has quantum numbers T_3= -$^1/_2$ and Y=1, so the e_L e_R “oscillation” is as if e_L gives a “charge” T_3=-$^1/_2$ and Y=1 to the condensate to become e_R (T_3 =0, Y=-2) and the other way around.

Using one set of possible names for these components might be [Tanedo]: The electron and the “anti-positron” (also called e_R) are constantly switching identities back and forth (both have charge -1 but the e is L-handed while the anti-positron is chiral R and cannot interact with a W. The physical electron is a mixture of these two alternating components. The positron and anti-electron particles switch back and forth (both have charge +1 but the anti-electron is R this time and can interact with W while the “positron” is L and cannot.

Electron: e_L, left-chiral, charge -1, can interact with the W, Y= -1, T_3 = - ½ .
Anti-electron: p_R, right-chiral, charge +1, can interact with the W
Positron: p_L , left-chiral, charge +1, cannot interact with the W
Anti-positron: e_R, right-chiral, charge -1, cannot interact with the W, Y = -2.

In “Zig-Zag” oscillation, the electron e_L gives one unit of hypercharge Y to the Higgs Vacuum expectation value, vev (246 GeV), and becomes e_R (with Y = -1 → -2). Then e_R retrieves one unit of Y from the vev → e_L (Y = -2 → Y= -1 again). The electron charge Q = -1e stays constant because weak isospin, T_3, compensates for each change in hypercharge.

“Within the electroweak theory, there isn’t an electroweak force, there are always multiple forces at every stage.” The Higgs field “rearranges the weak-isospin and hypercharge forces, making the photon out of a mixture of the W^3 and B, the Z^0 out of a different mixture of the W^3 and B” gauge fields [Strassler]. When we say $U(1)_Y \times SU(2)_L$, we are thinking of the B as hypercharge (Y) and the W’s as weak isospin (T).

Quantum Mechanics emerged in an early sub-picosecond ultra-dense period of the Universe when everything was up close and personal (essentially no distance separations). All interactions were at the speed of light—all connections were light-like – but that doesn’t mean light as we think we know it. The photon epoch began at 10 seconds after antimatter annihilation. But the emergence of the first photons was at EWSB at 10^{-12} seconds (ps). Prior to that time, there were no photons nor any massive particles. Mass is a sub-light slow down due to zig-zagging. So, did quantum mechanics exist then? That regime hasn’t yet been tested.

Conclusions:

$E = \hbar\omega$ says that the mass/energy of an electron or any other massive particle is intrinsically specified as a “vibration.” *{ With spherical symmetry, perhaps this can be envisioned as a broad “s-wave” sharing the space of the “particle”}.* It could involve a vibration of each individual type of quantum field but is more broadly a superimposed generic law of Nature covering all cases including photons. Lattice phonons can also be considered to obey de Broglie relations except that phonons are “quasi-particles,” momentum is specially defined “crystal momentum,” energy uses periodic boundary conditions, and relativity doesn’t apply. Energy transfer is quantized and energy levels are harmonic oscillator levels.

Quantum vibration in general is quantum and not classical, and one difference for the quantum world might be an intrinsic use of complex numbers similar to use of quaternions for spins and hypercomplex Clifford algebras for QED, for EW “forces” and gluon QCD. The mathematician Michael “Atiyah felt that the four division algebras – real and complex numbers, quaternions and octonians – provided essentially the only mathematical natural way to account for the number of fundamental forces (four) or the number of generations (three) in the standard model" [arXiv:1910.10630)] .

Topics in modern physics may add new perspectives such as Higgs field zig-zags. de-Broglie waves might also be primarily information waves such that mass-energy is coded and decoded as identical to the concentration of vibrations in time, and momentum is the linear density of wavelengths in space.

Energy is King, and the rules of its quantization and conservation are imposed from on high; and energy transfer is in bundles of $n\hbar\omega$. "All quantum states comprise two physical components: one is the source of the energy (radiation, lattice vibrations, particles) and the other is the energy state" [Street]. The substance that vibrates is not anything familiar. We can use the term "energy wave," but it is not yet energy.

The resulting picture I most like corresponds to the "de Broglie double solution" where a particle is a wave-formed "soliton" with a frequency matching (and probably causing) the wave of the ψ-field. This picture came before any quantum-mechanics but was so challenging that it was quickly simplified into what is now called "de Broglie-Bohm" theory ("dBB"). The soliton became just a particle position (a "hidden" variable) "guided by a Schrodinger wave, ψ(x,t)." de Broglie wasn't able to complete this ideal because it necessitates use of non-linear equation; and the idea of a soliton did not yet exist. But its development is still encouraged [Collin].

Now whether people like it or not, dBB is functionally equivalent to orthodox quantum mechanics but has a very different interpretation – it works and gives the same answers. It is not forbidden because it is a "non-local" hidden-variables theory – and we now know that ordinary QM is also "non-local" (correlations violate Bell inequalities). I would like to see some transactional "back-and-forth-in-time" physics added on to explain non-locality. People often don't like that a particle is a point with a well defined position – but the vibrating soliton smoothes that out. It is consistent with the latest mantra, "there are no particles, everything is fields."

Appendix:

Ontology is "the branch of metaphysics dealing with the nature of being: what entities exist and what are their relationships within a hierarchy. It is concerned with things, events, properties and facts about the reality of what is there. The philosophical views of physicists have evolved with developments in physics. Relativity makes topics like the coexistence of objects frame-relative and favors events over things. "Quantum mechanics could jeopardize both an ontology of events and an ontology of things." Physics used to be about ontology but then became positivistic based on observation. But we can ask, Observation, information, measurement and data – About What? "What could be the primitive ontology which could give rise to the appearances which our senses collect [Durr]." Einstein said, "It is the theory that decides what one can observe." And, "The supreme task of physics is to arrive at those universal elementary laws from which the Weltbild can be built up by pure deduction. There is no logical path to these laws; only intuition."

The major argument about interpretations of quantum mechanics has been whether the wave function is ontological or epistemological (also a question posed to SETI for aliens to answer -- perhaps in jest). Epistemology is the philosophical field concerned with the questions, "What do we know?" and "How do we know it?" The wave function would only represent our knowledge of what happens as observers. The word “real” is poor because it is presumed to mean “classically real.” A sub-quantum reality would be something else (a different ontology).

References

a. [Gurtler] R Gurtler and D.Hestenes, “Consistency in the formulation of the Dirac, Pauli, and Schrödinger theories,” J. Math.Phys, 16 573-584, 1975 , or geocalc.clas.asu.edu/pdf-preAdobe8/Consistency.pdf

b. [Tanedo] Flip Tanedo, https://www.quantumdiaries.org/2011/06/19/helicity-chirality-mass-and-the-higgs/ Famous chartoons for elementary particles.

c. [Penrose] Roger Penrose, The Road to Reality (A complete guide to the laws of the universe), Knopf, 2004, Chaper 25

d. [Jones] R. Michael Jones, Quantum Gravity and Mach’s Principle, book of Essays 1967 to 2019. In chapter 87 (p 572), “Real” waves are beyond QM. “An electron is waves of what?” (Question from Mike Jones) http://cires.colorado.edu/~mjones/physics/Quantum%20Gravity%20and%20Machs%20Principle%202019.pdf

e. [Peterson] David L. Peterson, “Reality of Schrodinger Ψ-Waves?” 9/14/18

f. [DP_spin] David L. Peterson, in A Stroll Through Physics , Book 1, 2014. www.Sackett.net/DP_Stroll.pdf ,

g. [Durr] Detlef Dürr, “Why Physics needs Ontology,” 2015, https://www.mathematik.uni-muenchen.de/~bohmmech/BohmHome/files/duerr_ontology.pdf

h. [Arndt] Y. Yaakov, M. Arndt, et.al., “Quantum superposition of molecules beyond 25 kDa,” Nature Physics 23 September 2019, {~ 23 TeV , λ ~53 fm ! }, https://www.nature.com/articles/s41567-019-0663-9?utm_source=commission_junction&utm_medium=affiliate { 2-m-long Talbot–Lau interferometer ,λ ~ 53 fm, five orders of magnitude smaller than the diameter of the molecules themselves.}

i. [Colin] Samuel Colin, Thomas Durt, and Ralph Willox, “de Broglie’s double solution program: 90 years later,” https://arxiv.org/pdf/1703.06158.pdf

j. Ruth Kastner, Understanding our Unseen Reality, Solving Quantum Riddles, 2015. {My Review of this book is at: https://www.amazon.com/gp/customer-reviews/R570ABO4X9NC3/ref=cm_cr_arp_d_rvw_ttl?ie=UTF8&ASIN=1783266465 }.

k. [Shuler] Robert L. Shuler, https://www.hindawi.com/journals/physri/2015/895134/ Physics Research International Volume 2015, Article ID 895134, 8 pages, “Common Pedagogical Issues with De Broglie Waves: Moving Double Slits, Composite Mass, and Clock Synchronization.”

l. [Street] R. A. Street, “Physics of quantum mechanics -- why are there phonons?” https://arxiv.org/pdf/1910.10020.pdf, Sept. 2019

m. [Strassler_Math] Matt Strassler Blog on Higgs interactions: Four Key Posts: **How the Higgs Field Works: with Math.** #4: https://profmattstrassler.com/articles-and-posts/particle-physics-basics/how-the-higgs-field-works-with-math/4-why-the-higgs-field-is-necessary/ #3 https://profmattstrassler.com/articles-and-posts/particle-physics-basics/how-the-higgs-field-works-with-math/3-how-the-higgs-particle-arises/ #2 https://profmattstrassler.com/articles-and-posts/particle-physics-basics/how-the-higgs-field-works-with-math/2-why-the-higgs-field-is-non-zero-on-average/ #1 https://profmattstrassler.com/articles-and-posts/particle-physics-basics/how-the-higgs-field-works-with-math/1-the-basic-idea/

n. [Strassler_Zero] **The Known Particles — If The Higgs Field Were Zero**, https://profmattstrassler.com/articles-and-posts/particle-physics-basics/the-known-apparently-elementary-particles/the-known-particles-if-the-higgs-field-were-zero/ Illuminating Figures of Known Particles.

Concrete Hidden Variables __ Rev. 3

Dave Peterson, 11/27/16- 12/6/16, for Cosmo:

Overview: We have been told that Bell inequalities and Bell tests for EPR experiments imply that no local (or classical) hidden variable (HV) explanations can suffice for the "unexpected" correlations between two EPR detectors in actual quantum mechanical (QM) experiments. Desiring an intuitive model that could work, we might suppose it to be due to pre-determined polarizations emitted from a source (pre-established coordination). And, without familiarity with concrete examples of what a hidden variable λ can be, we might weakly retain this idea in the back of our minds. Indeed one HV example included here is fairly close to the correlation for actual quantum mechanics (HV-A, see **Fig. A and Fig. C**). Surprisingly, a recent survey on the beliefs of physicists [10] showed that a third still say that physical properties exist prior to and independent of measurement, and only a third say that HV's are impossible. But a third also have general ignorance of Bell tests.

John Bell presented his revolutionary "Bell Inequality" for the Einstein-Rosen-Podolsky (EPR) entanglement paradox in 1964. Using essentially classical arguments, he showed that "any physical theory that assumes local realism cannot also predict all of the results of quantum mechanics [1]." He did this by introducing hidden variables represented abstractly by the symbol "lambda" and derived special inequalities that would be violated by actual experiments for entangled EPR particles. This involves performing one experiment with a pair of set angles (a,b) and then another with a different set angle, c, and then comparing them. The first Bell test looked something like this: $Correlation\ C(a,c) - C(b,a)\ -\ C(b,c)\ \leq 1$

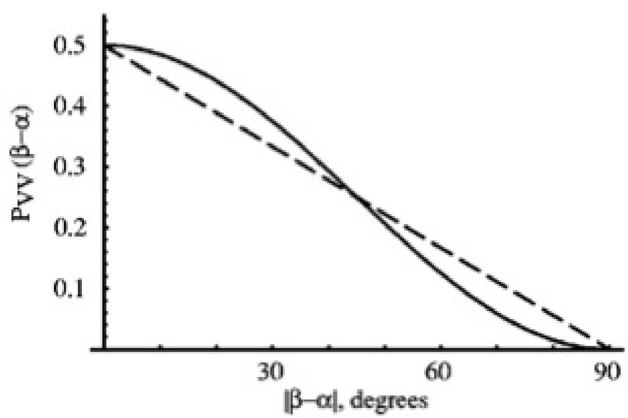


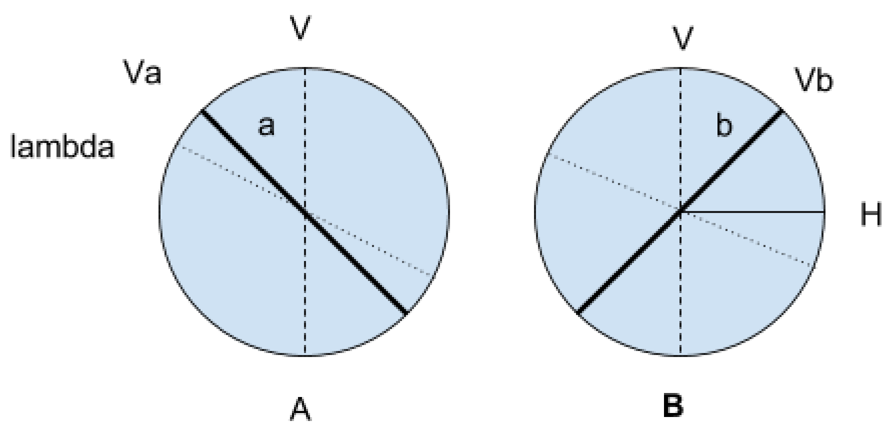
Fig. 4. Predicted polarization correlations for a quantum mechanical entangled state (solid curve) and a hidden-variable theory (dashed line).

<u>Figure A</u>: A standard hidden variable comparison for probability P(V V) versus the difference in angle of detectors (dashed line). Quantum Mechanics [QM] cosine curve (solid line) violates this at angles near 20 or 70 degrees (Figure taken from reference [2]).

<u>Examples shown below include</u>: 1) A derivation of the <u>standard</u> quantum mechanical correlation $P(V_a V_b) = \frac{1}{2} cos^2 (b-a)$. 2) The local hidden variable <u>"Triangle Plot"</u> **HV-A** Fig.A, (3) The interesting QM result using $\lambda's$ with Malus' law projections in **HV-B**. (4) A more intuitive LHV in **HV-C** (but still not as good as HV A), (5) And further discussions.

An example of one of Bell's arguments is available in a recent paper for our Boulder Cosmology group [1]. There are now many different types of test inequalities (e.g., "CSCH"), but they are all still called Bell inequalities. Well-tested experimental violations of Bell inequalities show that <u>all</u> local hidden variable approaches are doomed as a class. This was a major advantage of having a general abstract derivation. But, to really understand it intuitively, we need to show some concrete plausible examples of what a hidden variable might be.This paper largely avoids Bell inequalities and instead focuses on continuous graphs (like Figure A) for QM versus local hidden variable mathematics over all possible difference angles.

The initial proposed theoretical setup considered two oppositely directed spin polarizations from a central singlet state having total angular momentum zero. Particles are directed to two different spacelike separated detectors labeled A and B for spin measurement orientations labeled a and b. For spin, this might be Stern-Gerlach magnets with different north-to-south rotation angles (a and b). It is hard to actually do these spin angular momentum experiments; and it was found that use of photon polarizations was much more practical.



Polarizer Tilt Angles for photon detection.

<u>Figure B</u>.

Almost all experimental tests to date have been done using photons with a polarization basis that can be labeled as horizontal and vertical, H and V [see Fig. **B**], and considering just two entangled photons and two polarization detectors for the early experiments. For our convenience here (which is allowed) suppose one detector (say A to the left) has untilted xy polarizer axes (angle a = 0 = vertical V). Bell's key new idea was to consider rotating the other detector by an angle that is **not** the traditional b= 0, 45, 90, 135 or 180 degrees but rather some angle in-between (e.g., see graph differences in Fig. A). Initial experiments used sequential two photon decays from atoms in an atomic beam (e.g., excited Ca-40 decaying to green and violet entangled photons [6]). But now it much more convenient to get entangled photons from laser beams passing through non-linear crystals (e.g., nonlinear crystal Type-1 down-conversion sources so that the two output photons both begin with the same polarization HH or VV or their superposition. To process this for detectors A and B using a hidden variable, λ, we need to specify: Its name, its probability distribution $\varrho(\lambda)$ [*which does not have to be uniform*] , explicit functions for individual measurements $A(\lambda, a)$ *and* $B(\lambda, b)$ [*e.g.,* $A = A(a-\lambda)$] or how they are going to be used (what are the rules).

We require: $<A> \; = \int \varrho(\lambda) d\lambda \, A(\lambda, \alpha)$ and $<B> \; = \int \varrho(\lambda) d\lambda \, B(\lambda, b)$. And then for measured outputs using intuitive hidden variable, λ, and joint settings a and b: we need to integrate [1].

$$P(a,b) = \int \varrho(\lambda) A(a,\lambda) B(b,\lambda) d\lambda, \; where \; 0 \le \varrho(\lambda) \le 1 \; and \; \int \varrho(\lambda) d\lambda = 1. \; EQN \; 1$$

This is the standard LHV form. Notice here that the separate LHV detectors A and B only depend on their local settings a, b, and λ .For averages, we sweep through an ensemble of hidden variable values according to a probability density $\varrho(\lambda)$. The hidden variable here is: a pre-determined input photon polarization angle $= \lambda$ [very light lines in Fig. B] and use of sine or cosine electric vector projections onto polarizer axes.

What makes the actual quantum mechanical case **different** from intuition is that polarization is not defined prior to measurement (passing through the polarizer to a detector). But as soon as one photon is detected the other instantaneously (non-locally) "is projected into a state of polarization parallel to" the first result (V *or* H , e.g., see Aspect [5]) . Whether this is a photon on A or on B is totally random (collapse is random). I call this reduction "**SNAP-TO"** (as in a soldier snapping to attention or a computer visual "snap to grid"). This is different from the intuitive but naïve idea that perhaps the photons were initially tilted at the same angle from the source and kept that alignment up to the time of detection (called "real"). For that case, the initial hidden variable angle could be anywhere from 0-180 degrees (i.e.,0- π radians), a uniform distribution [$\varrho(\lambda) = cnst. = 1/\pi$ radians, so that $\int \varrho(\lambda) d\lambda = \pi(1/\pi) = 1, \; \lambda \in [-\pi/2, +\pi/2]$] -- an ensemble of all possible predetermined polarization angles. This provides a concrete example where the hidden variable lambda is merely any predetermined tilt angle for both of the photons per event.

QM Actual Quantum Mechanics Calculation Let's begin by first looking at the actual physical **QM** calculation for the coincidence of detector hits for vertical polarizer settings [2]. Begin with a left polarizer A having angle a = V = "|" and right polarizer B having angle b and V_b "/" and look at probability coincidence P(VV) meaning V's being vertical in the tilt angle bases of their respective polarizers [Fig. B]. Let the initial polarization state of two entangled photons be given in a neutral untilted basis: Eqn. 2

$|\psi_{EPR}> \; = (1/\sqrt{2}\,[\,|V>|V> \; + \; |H>|H>]$, *so* $|V_a> \; = \cos a|V> \; - \sin a|H>$ *and*

$|H_a> \; = \sin a\,|V> \; + \cos a\,|H>$. *Let* $|\psi_{DC}> \; = \cos\theta|H_1>|H_2> \; + \sin\theta\,|V_1>|V_2>$,

"DC" means down conversion, entangled photons are 1 and 2 on the untilted vertical (y axis) and horizontal (x axis). And we make the initial laser beam entering the nonlinear crystals for down conversion to have polarization at $\theta = 45^o$ *so that* $\cos\theta = \sin\theta = 1/\sqrt{2}$. Then $|\psi>_{DC} = |\psi>_{EPR.}$ There are four basic types of 'Bell states", but we will only use this one. Then project the down converted state onto tilted-vertical polarizations as in Fig. B. Then

$$P(V\,V) \Rightarrow P(V_a V_b) \; = \; | < V_a| < V_b|\,|\psi_{DC}>|^2 = (\; 1/\sqrt{2}\,)^2 \; |\sin a \sin b \; + \cos a \cos b|\,)^2$$

but $\cos(a-b) \; = \; \cos a \cos b + \; \sin a \sin b$, *so* $P(V\,V) \; = \cos^2(a \; - \; b)\,/2$ ANS.

Or, if we set angle a at 0 and replace angle b by the difference angle $\alpha = \; b - a$,

$$P(V\,V) \Rightarrow P(V_o V_\alpha) \; = \; | < V_o| < V_\alpha \; | \; |\psi_{DC}>|^2 = (\; 1/\sqrt{2}\,)^2 \; |\sin 0 \sin \alpha \; + \cos 0 \cos \alpha|\,)^2$$

$= \; \cos^2\alpha\,/2$. (Again). We can rewrite this also as:

$$P(V\,V) \; = \; (1/2)\cos^2(b-a) \; = (1/4)[1 + \cos 2(b-a)]. \quad EQN\ 3$$

The resulting parameter used for actual quantum mechanics tests is solely the difference in tilt angles of the two detectors [say angle alpha = $\alpha = \; (b \; - \; a)$]. QM calculations result in a Bell correlation depending on cos(2α) [note: for fermion electrons it would be just -cos(α)]. Suppose again, by rotational symmetry and convenience, that the left device A-angle is vertical, a =0= "| "; and there is a particular lambda angle $\lambda \sim$ " \", and right detector tilt angle may be α = "/ ". For the quantum case, a hit on the vertical detector "|" snaps the other photon also to "|" so that the only relevant polarizer angle to project onto is alpha for the second detector, B.

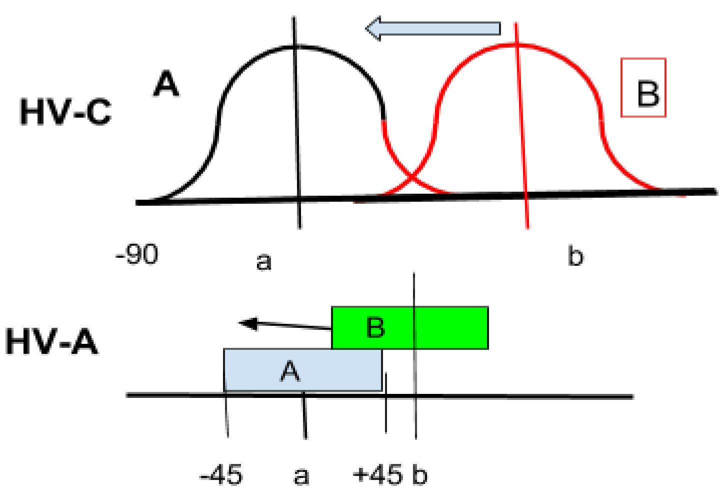


Figure C: Correlation Integrand shapes for overlap of A and B for Convolution.

Example HV-A: 2 joint snap-to's without Malus' law, **The Triangle Plot** (Fig. A)

Now, what about those plots we've seen before of "naïve" triangle (capital Lambda shape or "tent map") approximating the quantum cosine curve. Straight slanted lines go from the top of the cosine curve to its bottom valley – a not too bad approximation. Well, this case is for an implementation of hidden variable density using step-functions. This first concrete example is also the most popular example. Instead of vector electric projections to the polarization axes, we snap-to when within a certain domain of angles. This rule is that If angle HV $\lambda =$ " \" lies within 45^o of $\alpha =$ "/" and 0 = "|", then they both registers as V ="|" [e.g., ref. 2]. Here we are not using sine or cosine projections. In electrical engineering (EE), this $\varrho(\lambda)$ distribution is called a "rectangular" or Pi= Π shape. The probability $density\ \varrho(\lambda) \ = \ h\Pi(2\lambda/\pi) \ = height\ \ h = \ 2/\pi \ for\,|\,\lambda\,| \ \le \pi/4 \ = \ 45^o,\ (else\ 0).$

$$Then,\ \int_{-\pi/4}^{+\pi/4} \varrho(\lambda)d\lambda \ = \int_{-\pi/4}^{+\pi/4} h\Pi(2\lambda/\pi)d\lambda \ = \ 1.$$

The domain width of lambda is $w = [\pi/4 \ - (-\,\pi/4)] \ = \pi/2$. The special EE shape functions $\Pi(x)\ and\ \Lambda(x)$ are both understood to have unit height, and the form $h\Pi(\,(\lambda - a)/b)$ is centered at a, has total width b and height h. **We require λ to be close to <u>both</u> the left and right detector polarization** (A, B); and we care about the overlap correlation (angular overlap width = W radians) Figure C.

This is exactly the problem shown by animation in Wikipedia [3] for convolution that outputs a triangle shape (please look, it is kinda neat! See Fig. C). Convolution calculations can be hard and often require numerical methods, but ordinary calculus can also be used [e.g., 7]. This case with uniform distributions is much easier: there are no curves or slopes in these functions, so we can just apply simple geometrical thinking. Since we've conveniently chosen detector A to always have a vertical orientation (a = zero tilt from vertical), we can position the lambda density also at rotation zero leaving us with detector B with rotation alpha (the difference from b minus a angles). Overlap enables both a and b to be in range of $\lambda = \ \pm\,\pi/4$.

$$P(V\,V) \ = \ \int \varrho(\lambda)\,A(V\,,\lambda = 0)\,B(\alpha,\ \lambda)d\lambda \ = \ \int \varrho(\lambda - 0)\ \Pi(2[\lambda - 0]/\pi)\ \Pi(2[\lambda - \alpha]/\pi)\ d\lambda$$

This has the correct LHV form of Eqn. 1. And we also normalize the Maximum probability (when full rectangles overlap) to be N= 0.5 when the difference angle $\alpha = \ 0\ .\,So\ P \ = h(Nh)(W = w) \ = \ (2/\pi)(0.5)(\pi/2) \ = 0.5$ (half because we are ignoring another output that could have been P(HH)). There may be two domains to consider for overlap width W = right overlap minus left of overlap domain: [Slope Functions, Eqns A1]

$$\alpha > \ 0 \ \Longrightarrow W(\alpha) \ = \ \pi/2 \ - \alpha,\ and\,\alpha < \ 0 \ \Longrightarrow W(\alpha) \ = \ \pi/2 \ + \ \alpha,\ so$$

$$P(V\,V) = (h)(1)W \ = \ 1/2 - \alpha/\pi.$$

And these are just the equation for the triangle (Fig. A) with left slope up and right slope down ending at $\pm\,\pi/2$ [Figure A]. For the fermion-electron case, the slopes would end at π -- twice as wide. Bell's theorem says that it is impossible to find any local hidden variable example that can give actual quantum mechanics results. So Example HV A modestly fails to agree with QM. But, this seemingly kludgy "double snap" ends up being much better than some other attempts with local hidden variables (such as HV C below).

Example HV-B. Simple Malus' Law projections of pre-existing photon polarization (angle λ onto two analyzer orientations and b or 0 and $\alpha = (b - a)$. This sounds like a hidden variable calculation and lambda might really project initially onto the polarizers. But then lambda gets discarded. Overall, this is a little strange.

A detection on the "|" detector here leaves the other photon where it was at $\lambda =$ "\" (no snap-to) so that the electric vector now has to project to the difference in angles of $\alpha - \lambda =$"/ - \" . [We are assuming Malus' Law for real electric fields of a photon projecting onto the polarizer angle for detector A and separately for detector B]. Classical calculations may then require an integration over the ensemble of all local "hidden variable" λ angles.

Compare, QM versus use of HV= lambda and trig projections onto two polarizer angles:

$$P(VV) \Rightarrow P(V_o V_\alpha) = \int_{-\pi/4}^{+\pi/4} |<V_o|<V_\alpha| \ (|V_\lambda>|V_\lambda> + |H_\lambda>|H_\lambda>|^2/2 \cdot \varrho(\lambda)d\lambda =$$

$$= ((1/2)\int[<V_o|V_\lambda>\cdot<V_\alpha|V_\lambda> + <V_o|H_\lambda>\cdot<V_\alpha|H_\lambda>]^2 \varrho(\lambda)d\lambda =$$

$$= (1/2)\int(cos(0-\lambda)\, cos\,(\alpha-\lambda) + sin(0-\lambda)\, sin(\alpha-\lambda)\,)^2 \ \varrho(\lambda)d\lambda \qquad \leftarrow \ Eqn.\, B1.$$

$$=(½)\int \ [sin^2\lambda\, sin^2(\alpha-\lambda) + cos^2\lambda\, cos^2(\alpha-\lambda) - 2sin\,\lambda\, sin\,(\alpha-\lambda)\, cos\,\lambda\, cos\,(\alpha-\lambda)] \ \varrho(\lambda)d\lambda$$

$$\varrho(\lambda) = (\,height\ \ h = 1/\pi)\ for\,|\,\lambda\,| \le \pi/2 = 90^o\ (else\ 0). \quad Then, \int_{-\pi/2}^{+\pi/2} \varrho(\lambda)d\lambda = 1.$$

At first, this approach might be another Convolution Integral. Goal, for each alpha, evaluate the integral for lambda and then sweep through possible alphas for a final plot of P(VV) versus the polarizer difference settings alpha.

A peak result occurs for alpha = 0:

$$P(VV)(@\alpha = 0) = (0.5)\int(cos^2\lambda + sin^2\lambda)^2 \ \varrho(\lambda)d\lambda = 0.5\int 1\ \varrho(\lambda)d\lambda = 0.5\,(1) = 0.5.$$

And for $\alpha = 45^o$, $integrand\ A1 = cos\,\lambda((1/\sqrt{2})(cos\,\lambda + sin\lambda) - \ sin\lambda((1/\sqrt{2})(cos\lambda - sin\,\lambda) =$

$$= (1/\sqrt{2})(cos^2\lambda + cos\lambda sin\lambda) - (1/\sqrt{2})(sin\lambda cos\lambda - sin^2\lambda) = (1/\sqrt{2})(1),\ P(VV) = 1/4$$

BUT, look more carefully at that integrand in Eqn. B1. Eqn. B2 :

$$cos\,a\,cos\,b + sin\,a\,sin\,b = cos(a-b) = cos(0-\lambda)cos(\alpha-\lambda) = cos(0-\lambda-(\alpha-\lambda) = cos(\alpha) \ !!$$

And then: $P(VV) = P(V_oV_\alpha) = cos^2\alpha/2.$ This result is just QM!!! ANS.

The lambda contribution subtracts away! [I only noticed this after doing a spreadsheet calculation]. There is no need for convolution over all lambdas. The lambda angle can be anything or everything.
This HV lambda pre-existing orientation works like QM

And, note that Eqn. B1 is **not in the LHV form of Eqn 1.** Contributions from A and B are mixed together (arguments with $(\alpha - \lambda)\ and\ (0\ -\ \lambda))$. But **Local** hidden variables require separating A and B in the HV equation.
Most HV equations begin with finding averages of results for tests A and B separately.

For example, $< B(v) >\ =\ \int\ \varrho(\lambda)d\lambda\ B(\lambda, v, \alpha)$ where v is a preparation direction (like H or V). If we let $B(\lambda, \alpha)\ =\ |< V_\alpha || V_\lambda >|^2,\ then\ B\ =\ cos^2(\alpha - \lambda)/2 = cnst.$ (very constructive) Eqn A3.

One might think that a proper approach should really include horizontals:

$B(\lambda, \alpha)\ =\ |< V_\alpha|\,(|V_\lambda>\ +|\,H_\lambda>)|^2\ =\ |cos(\alpha - \lambda)\ +\ sin(\alpha - \lambda)|^2$

$= |\sqrt{2}\ sin(\alpha - \lambda\ + \pi/4)|^2$. And it is not clear that this should be dismissed

HV-C: A more intuitive proper LOCAL hidden variable calculation using cosine-squared.
Somewhat like Eqn 3 above, begin with an optical Bell calculation for the individual detectors: Let

$$A(a, \lambda) = N cos^2(a - \lambda),\ \Rightarrow\ \ < A >\ =\ \int_{-\pi/2}^{+\pi/2} N cos^2(a - \lambda)\ d\lambda/\pi\ = (N/\pi)\ [(a - \lambda)/2\ +\ sin2\ (a - \lambda)/4]$$

evaluated at limits to get $< A >\ =\ N/2\ =\ 0.5N.$ $used\ \varrho(\lambda)\ =\ 1/\pi\ and$ the sine contribution drops out). [A "normalizer" N was added just in case we wish to modify all results at the end.(for example getting a better fit to the QM result using $N\ =\ \sqrt{2}\ \ --\ a\,fudge)]$. But we really should be using just N = 1.

So now we can evaluate the correlation of a and b using the standard LHV form Eqn. 1.

$$P(V_a V_b)\ =< AB >= \int_{-\pi/2}^{+\pi/2} (d\lambda/\pi)\ cos^2(a - \lambda)\ cos^2(b - \lambda)\ N^2\ =\ \int_{-\pi/2}^{+\pi/2} (d\lambda/\pi)[(cos(a - \lambda)cos(b - \lambda)]^2\ N^2$$

At first this form looks like another convolution is needed (Figure C). But it can also be done just using calculus.

Expand $(cos\ x\ cos\ y)^2\ =\ (0.5\ cos(x + y)\ +\ 0.5\ cos\ (x - y))^2,\ and\ use\ \int\ cos^2 z\ dz\ =\ z/2\ +\ sin\ 2z\ /4.$

The result is $P(V_a V_b)\ =< AB >\ =\ 1/8\ +\ cos^2\ (a\ - b)/4.\quad \neq\ QM\ :\ cos^2(a - b)/2$!

We might kludge this up by using $N^2 = 2\ \rightarrow\ < AB >\ =\ (1/4\ +\ cos^2(a\ -\ b)/2\ )$

But it is still poor because of the 1/8th or ¼ offset value. We actually did much better using what seemed to be silly rectangles in HV-A.

[As a check, for

$\alpha = 0\ (max),\ 1/8 + cos^2(0)/4 = 3/8,\ also = \int cos^4\lambda d\lambda = (3/\pi 8)(\pi/2 -- \pi/2) = 3/8.$

Non-Local:

For reference, the most popular "non-local hidden variable" is the de Broglie-Bohm "position" x(t) along with the velocity of a moving particle. Remember that Copenhagen doesn't believe in the existence of trajectories, but dBB ~ QM works [9]! It is equivalent to usual QM but in a different form and different interpretation. In Bohm theory, "the non-local correlations are a consequence of the non-local "quantum potential," which exerts suitable torque on the particles leading to experimental results compliant with quantum mechanics [8]." dBB is not very popular, but it was intended as just an example of non-local hidden variable theory -- and as a counter-example, it revealed an error in von Neumann's "proof" of no hidden variables.

A separate class of non-local hidden variables was introduced by **Leggett** in 2003 along with a new inequality for testing. Assumptions are 1: realism (pre-existing properties independent of measurement) e.g., polarization u for A and v for B, 2: "physical states are statistical mixtures of sub-ensembles with definite polarization where" 3: Malus' law cosine projections apply for each sub-ensemble. A new nonlocal parameter, η, is introduced for arranging measurement settings across space-like separation of detectors A and B (often called "Alice" and "Bob"). Large statistics averagings are arranged (or contrived) to satisfy some QM expectation values. The contrivance is complex, so as just a partial sketch: the distribution for $\lambda \in [0,1]$ is decomposed into two parts for A at value L and into 3 different parts for B. For example:

$A = A(a,u,\lambda) = +1\ for\ \lambda \in [0,L),\ -1\ for\ \lambda \in [L,1],\ L = 0.5(1 + u \cdot a),\ B = B(a,b,u,v,\lambda)$

That is, Bob does all the statistical contriving [8], and he knows about Alice's settings "outside of space-time".

Actual testing of Leggett versus QM result in plots somewhat like Fig A that differ significantly for tilt setting difference $\alpha \sim \pm 30^o$ showing that "non-signalling correlations" don't work.

Conclusions:

In the mathematics above, we have bypassed the vast subject of "Bell inequalities" tests. Instead, we have addressed the continuous graphs of QM correlations versus local hidden variables models as a function of two polarizer-detectors having tilt angle differences, $\alpha = (b - a)$ (e.g., Figure A with QM versus concrete example "HV-A" with a modestly good fit). We have derived the fundamental quantum mechanical correlation equation: $P(V\,V) = cos^2\alpha/2$, and we used this result to attempt a LHV for A and B as cosine-squares (HV-C, but it didn't work very well -- much worse than HV-A). Then we came up with an example

HV-B that unintentionally ended up being the same as QM. But thought revealed that it wasn't Local.

A great majority of current journal articles seem to solidly support the conclusion that Bell-theory implies **non-locality** ("spooky action at distance") and the impossibility of Local hidden variables (LHV's). Yet, a recent survey on physicist's beliefs express residual doubt [10] at about one-third of physicists. Additional beliefs from the survey suggest that 2/3rds believe that true randomness is inherent in QM detections, 2/3rds believe that we need to have interpretations of quantum mechanics; yet ¾'s believe either in Copenhagen or simply don't care (which is another aspect of Copenhagen).

References:

1. Bill Daniel, "EPR Paradox and Bell's Theorem," see: http://www.sackett.net/cosmology.htm for November 21, 2016 Cosmology Schedule.
2. Dietrich Dehlinger, "Entangled photons, nonlocality, and Bell inequalities in the undergraduate laboratory," Am. J. Phys 70(9), Sept. 2002, pg. 903].
3. Convolution, https://en.wikipedia.org/wiki/Convolution
4. "Bell" en.Wikipedia.org, Bell's Theorem, Local hidden variable theory, /wiki/Bell_test_experiments, en.wikiversity.org/wiki/Bell's_theorem/Inequality, and see Nick Herbert:http://quantumtantra.com/bell2.html.
5. Alan Aspect, "Bell's Theorem: The naive view of an experimentalist," in Arxiv 0402001
6. DP, "Test of Quantum Entanglement" --Aspect Experiment," 7/7/2011, on web at www.sackett.net/DP_Stroll.pdf .
7. Rodolfo E. Firpo, "A Method to Easily Visualize and Solve a Convolution Integral by Direct Integration," Aerospace Report No.TOR-2012(1301)-1October 27, 2011
8. A. Zeilinger, et. al., "An experimental test of non-local realism," ArXiv 0704.2529 [quant-ph] 6 Aug 2007.
9. Mike Towler, University of Cambridge, Bohmian Course (I've gone through it) http://www.tcm.phy.cam.ac.uk/~mdt26/pilot_waves.html An Introduction to Pilot-wave theory.
10. "Surveying the Attitudes of Physicists Concerning Foundational Issues of Quantum Mechanics," n=149, arXiv:1612.00676 [physics.hist-ph] 2 Dec 2016.

[Google doc on gmail Google drive]   davepeterson137@gmail.com

DeBroglie-Bohm Interpretation of Quantum Mechanics.

Dave Peterson, 4/28/10- 5/30/10.

There is actually a **course on Pilot Wave Theory (PWT)** at Cambridge [1].

I like this interpretation and suspect that it could also be tied to Cramer's transactional interpretation. Its lack of general acceptance may be historically due to Bohr's charisma, Heisenberg's antagonism, and Bohm's unusual personality and life (McCarthyism defeating American Marxism, studies in psychology, consciousness, Krishnamurti, fearfulness, depression). But physics should not depend on authority figures, it should be objective.

The key formulas of the deterministic Pilot Wave Interpretation of Quantum Mechanics from Louis De Broglie (1924-1927) and David Bohm (1952) are the 'usual' Schrodinger Equation: **1): $i\hbar\, \partial\psi/\partial t = H\,\psi = -\Sigma(\hbar^2/2m_i)\nabla_i^2\psi + V(x_1,\ldots.x_n)$** and the **"Guidance Equation"** **2): $m_i dX_i/dt = \nabla_i(\hbar\, \mathrm{Im}\ \ln\psi) = \nabla_i S(X_i,\ ,X_n,\ t)$**

where S = phase. This equation simply says that particles actually exist in non-relativistic quantum mechanics and have the properties of a continuously changing location, X with a velocity v = dX/ dt governed by the pilot wave psi, ψ. The X's are in a configuration space; the guiding wave propagates in multidimensional configuration space, and that is the source of nonlocality and entanglement. Note that just as ψ is not a classical field, the Bohmian particles are not classical particles—they are not bearers "of properties other than position" and velocity. Since deBB uses both ψ and X, 'complementarity' is not needed and 'measurement' is unnecessary. The horribly misleading term 'measurement' should be dropped anyway—use 'experiment' instead. Measurement pre-supposes that what is measured pre-existed prior to measurement. Equations 1) and 2) are Postulates in Bohm theory. A third postulate for $\rho = \psi^*\psi$ could be added (but some claim it can also be derived). Unlike Copenhagen, no other postulates are needed.

Bohm writes a polar form: **$\psi(x,t) = R(x,t)\exp(iS(x,t)/\hbar) = |\psi|\, e^{iS/\hbar}$.**
The velocity field is intuitively straightforward when the exponent is a plane wave: S = px-Et so that ∇S yields p: $\mathbf{v} = dX/dt \sim \hbar k/m = (\hbar/m)\,\mathrm{Im}(\nabla\psi/\psi) = (\hbar/m)\mathrm{Im}\,\nabla\ln\psi = \nabla S/m = j(x,t)/\,|\psi(x,t)^2\,|$ with j = probability current. There are coordinates (x_i) and also the positions of each particle in a system, X_i where positions are sometimes called "hidden variables" but are also often the measured result of an experiment (not so hidden). In "standard" (Copenhagen) QM, there are no particle trajectories or intermediate positions.

This is all that is needed to discuss Bohm philosophy. But if one desires causal details about the particle trajectories , then an additional complex Quantum Potential, Q, is also useful:
Newton's Law is $F = -\nabla U = ma = dp/dt = d(\nabla S)/dt$ [1]. So consider:
$\partial(\nabla_i S(x,t))/\partial t = \nabla_i\, \partial(\mathrm{Im}\ln\psi)/\partial t = \nabla_i\, \mathrm{Im}\,[(\hbar/\psi)\partial\psi/\partial t] = \nabla_i\,[(i/\psi)\{\hbar^2/2m_i\, \nabla_i^2\psi - V\,\psi\}\,] = -\nabla_i\,[\{\hbar^2/2m_i\, \nabla_i^2\psi - V\,\psi\}/\,\psi] = -\nabla_i\,[V + Q] = m_i\, a_i$. This suggests that one should set:

3). Quantum Potential $Q = -\,\Sigma\,(\hbar^2/2m_i)\nabla_i^2|\psi\,|\,/\,\,|\psi\,|$.

For stationary states (e.g., the ground state of the hydrogen atom), the velocity field and Q are independent of time and the velocity is zero! The quantum force $-\nabla Q$ cancels the classical force $-\nabla V$, and the particle is at rest. A particle between two impermeable walls also has a particle at rest. Take away the walls, and the particle will be accelerated to its expected speed, $v = \pm \hbar k/m$, by the quantum force $-\nabla Q$.

For the **two-slit experiment,** a particle actually passes through either the upper or lower slit and the usual psi wave goes through both slits. The particle trajectories from the quantum forces are highly non-intuitive and non-newtonian [8]. But a particle from the upper slit will only appear above the medial line on the detecting screen. The elaborate electron trajectories can be worked out. For massless photons, the curved trajectories can also be calculated and are different than for massive electrons [11].
In Bohmian mechanics, there is no need for a collapse postulate. The wave function of the total system (the universe) is described by the Schrodinger equation and never collapses. Collapse only pertains to the subsystem. Figures of the Quantum Potential away from the slits show flat mesas separated by short precipitous valleys. The derivative $F = -\nabla Q$ is zero for the usual detection peaks (semi-stable trajectories) but have kinks for the inbetween trajectories (rapid forces between classical peaks).

There now is actually a **course on Pilot Wave Theory (PWT)** at Cambridge which teaches this interpretation [1]. In deBB, "probability refers to the probability that a particle ***is*** at some position, rather than to its probability of being found there in a suitable measurement." PWT is just standard QM with this single semantic change in the meaning of a word. QM is a "dynamical theory of particle trajectories rather than a statistical theory of observation." It is claimed that all the interpretation problems raised in non-relativistic QM are essentially solved by the pilot-wave approach.

Feynman had "considerable respect" for Bohm; Hiley showed "how to obtain a Bohm approach in the momentum representation" [non-commutative quantum geometry", 2005]; and Bell was inspired to do his work after reading Bohm. Those who respect Bell should respect Bohm. The Feynman Path Integral integrates $L = T - V$ over all paths while PWT integrates $L = T - (V+Q)$ along precisely one path where amplitude curvature $Q = (\hbar^2/2m)\nabla^2 R / R$ is the quantum potential.
Psi ψ in PWT is a real physical field which influences particle trajectories.

A mistake by the founding fathers of QM was their insistence that familiar conserved Newtonian terms such as momentum, energy, and angular momentum still had meaning in the world of quantum mechanics [3]. Heisenberg said that defining a physical quantity means specifying how to measure it, but the act of measurement and the apparatus alters what might have been there. Bohm has real particle positions, but the trajectories are "strange" and non-newtonian. Taking m times v might be called 'momentum'—but it is no longer conserved. "The energy of the particle is strictly conserved only in special cases like stationary states. "Because the quantum potential can be large even when the quantum field is small, it follows that we will inevitably have non-conservation of both energy and momentum resulting from the fact that complete isolation of any quantum system is actually impossible." [7, pg 39]. The resulting structure of "standard" Copenhagen mechanics is strange, and standard quantum mechanics doesn't hold logically together. It is like the "Escher Waterfall" picture where each section looks ok but the whole is nonsense. [Similarly, in my old copy of Messiah from a class in 1966, I had drawn the “Devil’s pitchfork”—2 bars becoming 3 prongs—to show what I thought of its characterization of quantum mechanics].

There should not be any authority or dogma in physics. “A philosopher, engineer, mathematician, or chemist might accept the authority of the majority of physicists. But, if you are a physicist yourself, you are in the position to decide for yourself.” [3]. But first, there has to be awareness that there is any controversy at all—that fashion actually exists in physics more often than just rarely.
“The best current PWT models fields for bosons and particles for fermions. In all cases studied, the usual predictions are reproduced. One doesn’t need both the field and particle ontology for the same object” [1, Lecture 5]. In PWT, “True observables of the theory—the things that immediately present themselves in experiments --are the positions of particles, particularly that of the apparatus pointer.” “The ‘hidden variables’ are the observables…” [Lecture 4]. Rather than the old name of “Hidden Variables theory,” Bohm preferred “Causal Interpretation.” [7].
Decoherence is a diminution of interference, and is a concept introduced by Bohm. Decoherence does NOT solve the measurement problem in QM. The different branches still continue to exist but stop interfering. It doesn’t make the second Schrodinger cat go away. Note that Bohm theory is a “formulation” requiring calculations like the path integral formulation.

There are no instantaneous quantum jumps in PWT. Changes are a continuous process. And wave collapse and splitting universes are seen as “utter madness.” “Pilot Waves subsume quantum concepts of measure, complementarity, decoherence, and entanglement into mathematically precise guidance conditions on position variables. “ In contrast, some Copenhagen concepts are expressed just in words. The borderline between unitary development of the wave equation and its collapse from the quantum world into the classical world is vague.

The ‘particle in the box’ (x = -a to +a, lowest energy level) has $\psi = (1/2\sqrt{a})(e^{ikx} + e^{-ikx} = 2\cos(kx)) = \cos(kx)/\sqrt{a}$ where $k = 2\pi/\lambda = \pi/2a$. The fixed f(x) shape is due to interference between left and right moving waves. The polar form is $\boldsymbol{\psi = Re^{is/\hbar}}$ = $[\cos(kx)/\sqrt{a}]e^{-iEt/\hbar}$. Current $j = (i\hbar/2m)[\psi\nabla\psi^* - \psi^*\nabla\psi] = (\hbar/m)\mathrm{Im}\,(\psi^*\nabla\psi) = R^2\nabla S/m$. Then current $j = (\hbar/m)\mathrm{Im}(\text{real}) = 0$, and $\nabla S = \partial/\partial x\ (0 - iEt/\hbar) = 0$ too. So the particle velocity is zero! However, the supposed operator for KE is $p^2/2m \rightarrow -\hbar^2\nabla^2\psi/2m = (\hbar k)^2\psi/2m > 0$ – there is a disconnect between Bohm and Heisenberg views. Other stationary states like the Harmonic Oscillator [$\psi_n(x,t) = u_n(x)\, e^{-iE\, t/\hbar}$] would also have zero particle velocity.

The more general alternate form is $\psi = \sqrt{(2/L)}\,\sin(n\pi x/L)$ for x from 0 to L. This could also be written in complex form as $\psi = (-i/\sqrt{L})(e^{ikx} - e^{-ikx})$. The quantum potential is $\mathbf{Q(x) = (-\hbar^2/2m)\,\nabla^2 R/\,R}$ = $(+\hbar^2/2m)(k^2 = n^2\pi^2/L^2)$ so that Q is now what would have been the particle energy, $p^2/2m$. Again, the particle is not in motion. Q has no dependence on x so that force in the well = $-\nabla Q = 0$.

The hydrogen atom is slightly different: $\psi_n(r, \theta, \phi, t) = u_n(r,\theta,\phi)e^{-iE\, t/\hbar}$. Notice here that the Legendre Polynomials for θ could be expressed in complex notation, e.g., $P_2 = (1/2)(3\cos^2\theta - 1) = (3/4)\cos 2\theta + ¼$ where $\mathbf{\cos(2\theta) = (1/2)[e^{i2\theta} + e^{-i2\theta}]}$ from $-\pi$ to $+\pi$ radians. This again resembles a fixed $f(\theta)$ due to interference between forward and backward moving waves which cancels any net velocity. But, $\Phi(\phi) = e^{im\phi}$ does imply just positive velocity quantization—so here $v \neq 0$. The atom doesn't match the standard stationary form except when $m = 0$ – then the electron is at Bohmian rest. Here we have the phase exponential $\exp[i(m\,\phi - Et/\hbar)]$—which describes a "rotating plane wave." In this sense, it resembles the old Bohr atom orbits. In a special case of a superposition of 1s and 2s orbitals together, the electron particle vibrates radially with a radial period near $\tau \sim 4 \times 10^{-16}$ seconds.

PWT-course Lecture #6 showed Bohmian calculational methods like finite difference equations for trajectories. And PWT lecture #7 addressed, "Why does nobody like pilot-wave theory?" The answer is largely fashion. Bohm theory is called a strange alteration of standard quantum mechanics—but really Bohm theory is DeBroglie pilot wave theory which is the original quantum mechanics—not a new alteration of Copenhagen. In QM, philosophical discussions are necessary. Philosophical questions "diminish the dogmatic assurance which closes the mind against speculation" [Russell]. Alternative interpretations enlarge our conception of what is possible. Einstein and Schrodinger remained incredulous at stated certainties like, "In Quantum Mechanics there is no such concept as the path of a particle." [Landau and Lifschitz].

Like Copenhagen, Bohm theory doesn't reveal a physical mechanism for how nonlocality occurs. Other questions are: does the wave ψ carry diffuse mass, charge, and spin rather than the particle. Bohm theory suggests that spin lies in ψ rather than in the particle. Doesn't dipole-dipole force require random charged atomic dipoles? (rather than a fixed charge location). It wouldn't make sense for m, q, or s to be diluted by propagating everywhere.

The extension of PWT to quantum (non-classical) spin is accomplished by simply letting $\mathbf{-i\hbar\,\nabla_k \rightarrow \sigma_k \cdot(-i\hbar\nabla_k - e_k A(q_k,t)/c)}$ and $i\hbar\partial/\partial t \rightarrow i\hbar\partial/\partial t\; e_k\phi(q_k,t)$. Instead of $\psi^*\psi$ use $\psi^+\psi$ which is now the inner product over spinor space degrees of freedom. That is, use the Pauli Equation where $\psi(x,s) = \phi(x)\,(^a{}_b)$ – two component spinors.

Pauli: $\mathbf{E\psi = i\hbar\,\partial\psi/\partial t = [(p - eA/c)^2/2m - e\hbar/2mc\;\sigma\cdot B + e\phi]\psi}$. (Bjorken/Drell p13).

A criticism of Bohmian Mechanics is its believed lack of a relativistic form. "Several 'Bohm-like' models for relativistic quantum mechanics and quantum field theory do exist [9]. A good model should reproduce the predictions of QFT and include non-relativistic QM as a limiting case. In addition, the theory should possess 'beables' – clear ontology or entities with 'being.'
Bohm made a 'Bohm-Dirac' theory where $dX_k/dt = \psi^+\,\alpha_k\,\psi/\,\psi^+\psi$ where α is Dirac's alpha matrix. But it uses a 'common time for all particles' – not Lorentz covariant. This distinguished frame model is ok because it cannot be identified experimentally. Also there are "Field-beables for bosons and particle beables for fermions." Fermion number can vary, but its states are 'real.'

The Mach-Zender Interferometer examines wave interference after two beams merge from a rectangular path. The paths are constructed to initially provide constructive interference for one outcome direction and destructive interference for the other. This is usually facilitated by having two beam splitters (BS) using opposite orientations. Light hitting a semi-mirrored metal surface is reflected by π (half a wavelength phase change). If light passes through glass first before the mirror (the opposite orientation), it is reflected from the back side to air interface (a lower index of refraction) and hence has no reflected phase change. The destructive output then results from the phase change difference of ($\pi + \pi$) for one path minus π for the other . Penrose likes to refer to another case of i (90°) from each mirror resulting in $\Delta\phi = i^3 - i$ instead – but the result is the same.

Figure: Mach-Zender Interferometer:

```
   M                      | null path output '2' =>
   /----------------------/ -- → → output '1' constructive => wave detector.
   |                      | BS2        |ψ> = |ψr> + |ψt >
   | 'path r'             |
   |                      |    BS2 has reverse orientation from BS1
   |                      |
→ -/----------------------/ standard Mirror, M
 BS1      path 't'
```

Bohm trajectories follow probability current streamlines, and individual particles stick to these flux lines. The medial line of a two-slit interference experiment does not have crossing streamlines because of the symmetry of the currents. The above figure would become a particle detector if beam splitter BS_2 were removed. Then Copenhagen (CQM) will believe that outcome '2' comes from the transmitted lower path 't' and '1' from reflected path 'r'. Their particles plow straight through the BS2 region. **But Bohm QM will have no crossover of current or particle trajectories at the BS2 location** so that instead '2' comes from 'r' and '1' comes from 't'. This is one of the strongest differences in beliefs. CQM believes that pathways have inertia—continuity of momentum. BQM only sees $mv = \nabla S$ following flow without conservation of classical p.

Englert, Scully, Sussman, and Walther (ESSW, X. Naturforsch 47a, 1175, 1992) challenged Bohm trajectories using the incomplete Mach Zender Interferometer for particles (as a particle detector without BS2 present). There was considerable controversy about their claims over the next decade [12,13,14]. ESSW suggests a which way detector in path 't' at bottom (a small change in an n = 63 Rydberg Rb atom passing through a micromaser there). So one can measure the output particle energy and know which path was actually taken. Having a Bohmian particle detected at '2' after traveling the upper path 'r' but yet having a small loss of energy would be "surreal." Solutions include possible application of EPR, configuration paths which cross "over" each other or change from pure to mixed states due to the micromaser action. Mixed state incoherent wave function components can cross. Bohm theory is completely compatible with standard quantum mechanics and cannot provide any contradictions.

Side note on "Ether:" It is well known that in 1905 Einstein dismissed the ether. But it is generally not known at all that after 1916 he decided that earlier judgment was too radical. He really intended to just deny the concept of ether immobility or velocity. His new belief was that the general theory of relativity was incomprehensible without an ether—and he continued to refer to it in print for decades. His new ether was : **field $g_{\mu\nu}$ = ether.**

Empty space has physical properties—physical space is a "primary thing." The "old ether" was "inertial ether" and the new ether was "gravitational ether" or "total field" without any absolute motion. He wished it to also include electromagnetism and in 1924 ("Uber den Ather" – and later) particles also as states of space. In his "Mein Weltbilt" book from 1934 he says, "Physical space and the ether are only different terms for the same thing; fields are physical states of space." [Ludwik Kostro, QC173.6 I572, 1988].

In addition, not everyone is still convinced that Einstein SR is the best interpretation for special relativity [1,6]. "The pre-Einstein position of Lorentz and Poincare, Larmor and Fitzgerald was perfectly coherent as presented and is not inconsistent with relativity theory [J. Bell, 1986, "The Ghost in the Atom"—this is not well known]. Poincare had a preferred frame with 4-vectors and Lorentz invariance worked out also in 1905, but his work was eventually nearly forgotten. Interpretations are called: 1. the Einstein 3+1 space and time with equivalence of inertial frames, 2. Minkowski 4D with no past or future, and Lorentz 3+1 single ether preferred frame and absolute simultaneity – this one is preferred by the Bohmists. Lorentz was still using the term "aether" in 1911. Dirac in 1951 said, "we are rather forced to have an aether" [of course not the "luminiferous aether]. Lorentz allows the absolute instantaneous simultaneity needed for EPR correlations. It is possible to make a Lorentz QFT—QFT doesn't use rods or clocks or the speed of light but rather symmetries and Lorentz invariance of Lagrangian density. DeBroglie-Bohm seems to require a preferred reference frame in order to be relativized. [But a question is whether John Cramer may have an alternative explanation]. Einstein's early view was positivist—a philosophy that is no longer revered.

Cramer: [Quotation from a future space alien in one of his popular science fiction books: "Our science historians "have derived great amusement from your quantum mythology. They were particularly amused by your Copenhagen interpretation, with its state vectors that are altered by the thoughts of intelligent observers, and by your Everett -Wheeler interpretation with its splitting and resplitting into multiple universes. In this regard, your culture is unique among those that we have encountered. No other has provided such a remarkable demonstration of fertile creative desperation in seeking to understand physical behavior at the quantum level. We find these myths of yours quaint and charming."
["Einstein's Bridge," science fiction by physicist John Cramer, pg. 205].

John Cramer [4] notes that the Everett-Wheeler interpretation came out in 1957 but was ignored until Bryce DeWitt published an article about it in 1971. Now the theoretical physics community embraces it "in spite of its conspicuous inadequacy in dealing with the problem of nonlocality." Many believe that the Transactional Interpretation of QM (TI) only deals with photons. But trajectories of advanced and retarded waves do not have to be at 45°. The handshake for electrons will use "a negative charge with positive energy wave and a positive charge with negative energy wave, the latter reinterpreted as a positive energy electron according to the usual Dirac rules."
A primary perceived problem with TI is that "it is necessarily deterministic, requiring an Einsteinian block universe to pre-exist, because the future must be fixed in order to exert its influence on the past in a transactional handshake." [5]. "However, while block-universe determinism is consistent with TI, it is not required. A part of the future is emerging into a fixed local existence with each transaction, but the future is not determining the past, and the two are not locked together in a rigid embrace."

Matter Interference: In 2003, perfluorinated buckyballs ($C_{60}F_{48}$) became the "most massive single particles to display quantum interference." [M = 1632, 108 covalently bound atoms, size ~ 1 nm]. In a Talbot-Lau Interferometer, the interference decays as gas is added to the experimental vacuum [decoherence by 10^{-6} mbar pressure]. [Physics World 3/05--/18/3/5/1].

References:

1. **Mike Towler** Institute Workshop, Tuscany, **DeBroglie-Bohm quantum mechanics** { Pilot Waves, examples, measurement, nonlocality, relativity, calculations, status, metaphysics} ,: www.tcm.phy.cam.ac.uk/~mdt26/PWT/lectures/pilot_waves.html/bohm7.pdf , There are eight pdf lectures available at this site. Source found at "Not Even Wrong" by Peter Woit.

2. D. Durr, S. Goldstein, N. Zanghi, "Bohmian Mechanics as the Foundation of Quantum Mechanics." Nov. 13, 1995.
3. Roderich Tumulka, "Understanding Bohmian Mechanics: A Dialogue." Am. J. Phys 72 (2004) 1220-1226, www.mathematik.uni-muenchen.de/
4. Letter from John Cramer to Mike Towler, http://www.tcm.phy.cam.ac.uk/~mdt26/cramer_letter.txt

5. John Cramer book chapter 9. "The plane of the present and the new transactional paradigm of time." >1995 (details not recorded).
6. W. L. Craig, Quentin Smith, Einstein, Relativity and Absolute Simultaneity, Routledge, 2005.
7. D. Bohm, B.J. Hiley, The Undivided Universe, 1993.
8. Quantum Implications Essays in Honor of David Bohm, ed. B.J. Hiley and F. David Peat, Routledge, 1987.
9. Oliver Passon, "What you always wanted to know about Bohmian mechanics but were afraid to ask, Physics and Philosophy, 2006, https://eldorado.tu-dortmund.de/handle/2003/2318 or …/bitstream/2003/23108/1/003.pdf

10. Sheldon Goldstein, "Bohmian mechanics and the Quantum Revolution, arXiv:quant-ph/9512027v1 26Dec1995.
11. Photon double slit trajectory pictures, see Partha Ghose P.L. A 290 2001. physicsforums.com/showthread.php?t=31304/ and quant-ph 12Jan2009 Hrvoje Nikolic (Zagreb, Croatia) and also see arXiv:0904.2287 and arXiv:0811.1905. For both electrons and photons, trajectories never cross.
12. "Welcher Weg Experiments from the Bohm Perspective," B. J. Hiley (Web ~ 2005).
13. "What is Surrealistic about Bohm Trajectories? M.O. Terra Cunha (Brazil), arXiv:quant-ph/9809006v1 3Sep 1998.

14. "Do Bohm Trajectories always provide a trustworthy physical picture of particle motion?, Marlan O. Scully, Physica Scripta, Vol. T76, 41-46, 1998.
15. Partha Ghose, Majumdar, Guha, Sau, "Bohmian trajectories for photons." Phys Lett A 290, p 205-213, 11/01, or arXiv:quant-ph/0102071v2 11 Oct 2001.
16. New Scientist p 42, Muon g ~ 2, 22 May 2010, also in Chinese physics C, and hep-ph, and arxiv.org/abs/1001.4528. Experiment switched off in 2001—but it took a long time to study the data—and an improved experiment is planned at Fermilab ('P989' = Mu2e).

Questions:

1. Where in Bohm Postulates 1 & 2 is the configuration space? I think the answer is in the Σ over N particles—"Usual" QM just discusses one although entangled systems may discuss 2 or more. Psi and X also have all the particle locations or variables built into their definition too. Where is the 'measurement' apparatus in here—it should be part of the configuration, but that isn't carefully discussed in the papers I have. Perhaps this becomes clear in the calculational examples (no, not from the course). Most sources say that the non-local configuration space shows best in the quantum potential Q. Q is a function of the Configuration of N particles.

2. Where are the other particle properties besides position and velocity? Where is charge, spin, mass, momentum, energy? Is a particle without any of these standard particle properties still a particle. The particle still has the mass, but its energy and momentum vary with the quantum potential. Articles say that spin is contained in the modified wavefunction. All of this is being addressed.

Later Notes:

Present Bohmian trajectories for photons are difficult [Ghose, 2001] – a "10 x 10 dimensional representation describes spin-1 bosons." [Kemmer, 1930's, using special β matrices]. This provides a consistent relativistic QM for spin 0 and 1 with a conserved four-vector current [15]. The equation is $(i\hbar\, \beta_\mu \partial^\mu + m_o c\, \Gamma)\psi = 0$. For massive bosons, $v = c\,\beta$ and $\Gamma = 1$. For photons, Γ and β_i's contain Maxwell equations and
$v = c\, \psi^T \Gamma\beta\Gamma\, \psi / (\psi^T \Gamma\, \psi)$. The set of 2-slit photon trajectories resembles those for the case of electron interference (smooth streams for constructive peaks with kinks inbetween).

"The literal identification of eigenvalues with real physical quantities is the fundamental error in quantum measurement theory." [Lecture I].

Heisenberg's uncertainty principle 'undercut Schrodinger's premise that an electron's position and velocity could be simultaneously specified." But, in Bohm theory, mv is not a conserved momentum classically called 'p'.

"**Positivitism** was abandoned by American philosophers several decades ago, with physicists lagging one or two decades behind in this regard. It is hard to find American philosophers nowadays who will defend positivism, but it is not at all that difficult to find physicists mouthing positivistic slogans, particularly when in deep quantum mode." [letter from Shelly Goldstein to Steven Weinberg, 1996].

Virtual particles are considered only in perturbative expansions. This picture should not be taken literally! Exchange of virtual particles may not be real.

How is Bohm Theory different and unusual:

Having a particle position actually existing is not at all incompatible with having a particle velocity. Bohm theory is the simplest quantum mechanics and resolves or avoids the metaphysical paradoxes of Copenhagen QM. If history could be replayed and if Bohm theory had existed alongside deBroglie theory at a time just after the Schrodinger equation, the Copenhagen interpretation might never have existed and we would all be "Bohmists."

For standing waves (particle in a box, linear harmonic oscillator, S states of atoms), the Bohm particle is not moving—is also stationary. But for atomic orbits with m values > 0, the Bohm particle orbits like a Bohr particle.

Particles move like their probability fluid streamlines. Streamlines do not cross each other, and neither to Bohm particles. So, for the two slit interference, a particle going through the top slit stays at the top of the detecting screen. And this is also true for a Mach-Zender interferometer—the particle motion is the opposite of the mental picture accompanying usual QM.

There is fad and fashion in physics. After 70+ years of QM, it is not obvious that the right side won. But there is now much more openness about QM interpretations in general.

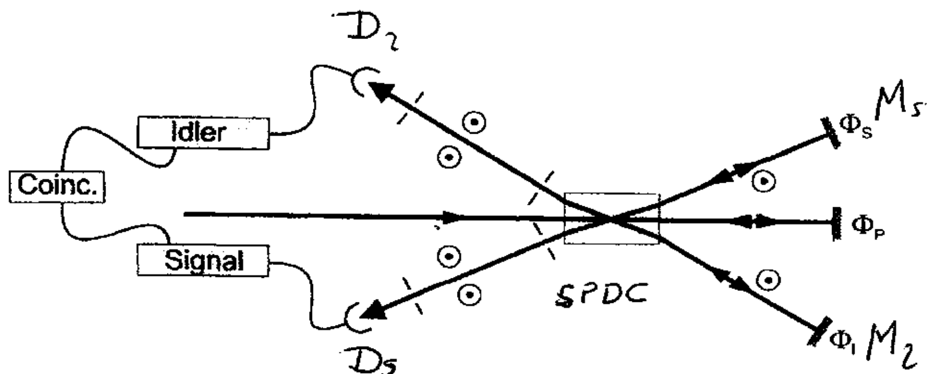
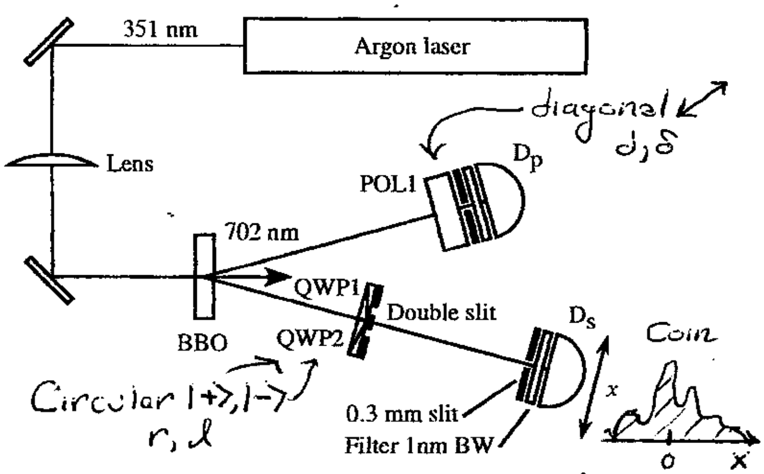
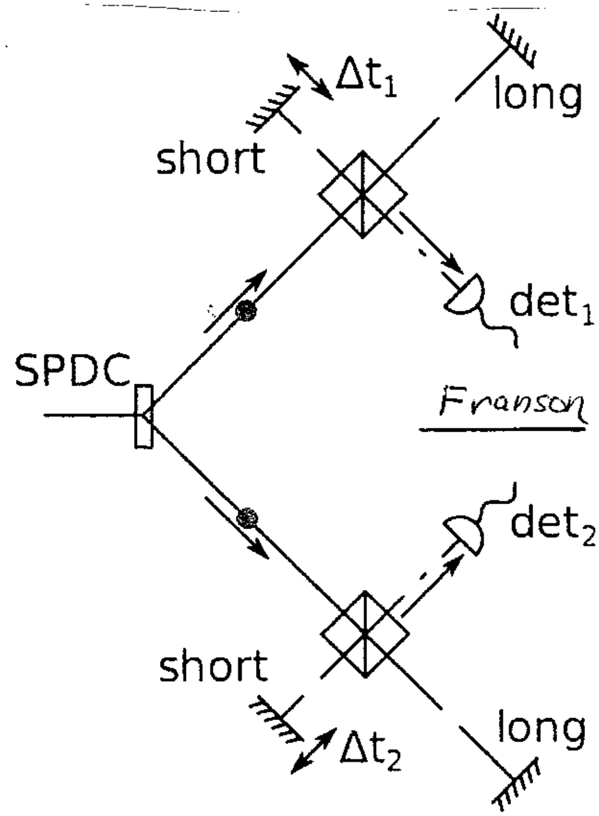
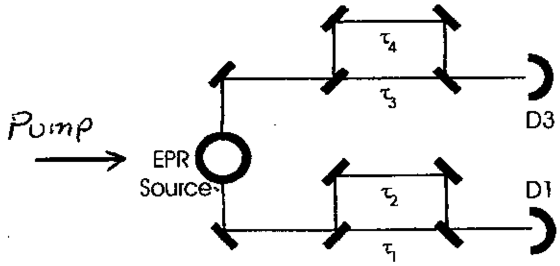


FIG. 1. Basic setup of the interferometer: A nonlinear crystal is pumped by a laser to produce photon pairs via parametric down-conversion. Directing the pump beam back through the crystal gives a second possibility to create the photon pair. Reflecting the pairs created in the first process back into the crystal makes them indistinguishable from pairs created in the second process, and interference occurs.



Walborn 2008, Double Slit Quantum Eraser

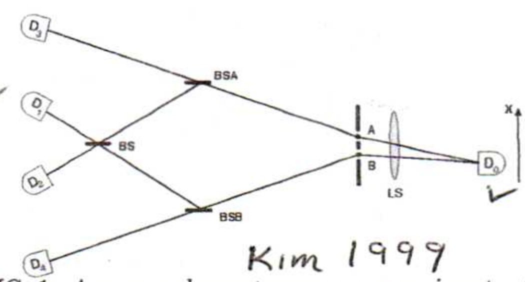




1989

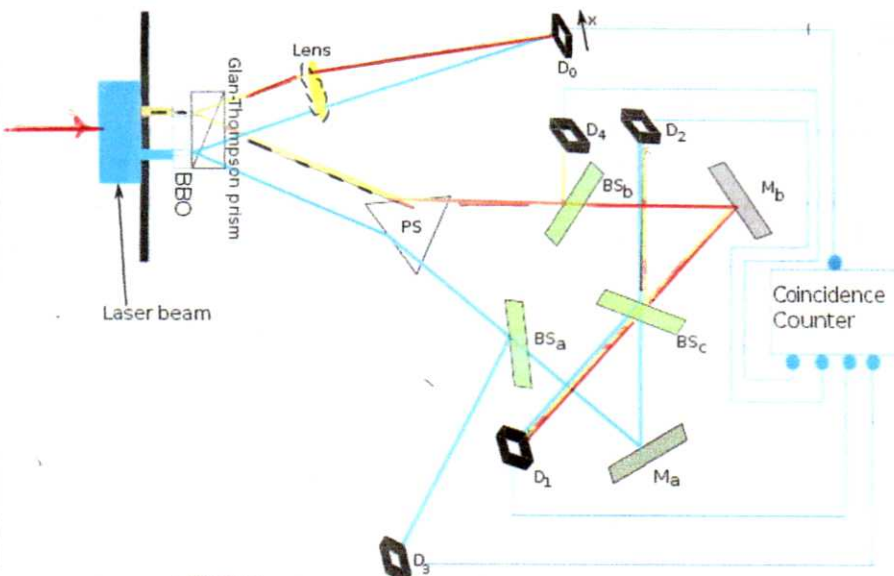
FIG. 1: Fig 1. A Franson interferometer. Two photons travel in opposite directions from an EPR source. Each photon then enters an imbalanced Mach-Zehnder interferometer, such that there is no single photon interference. The temporal pathlengths through the arms of the interferometers are denoted τ_i. Detectors in paths 1 and 3 measure the outputs of the interferometers and the coincidences between the detectors are recorded.

Entanglement Experiments



Kim 1999

FIG. 1. A proposed quantum eraser experiment. A pair of entangled photons is emitted from either atom A or atom B by atomic cascade decay. "Clicks" at D_3 or D_4 provide which-path information and "clicks" at D_1 or D_2 erase the which-path information.



1999

Figure 2. Setup of the delayed-choice quantum-eraser experiment of Kim et al. Detector D_0 is movable

FIG. 2. Schematic of the experimental setup. The pump laser beam of SPDC is divided by a double-slit and incident onto a BBO crystal at two regions A and B. A pair of signal-idler photons is generated either from A or B region. The detection time of the signal photon is $8ns$ earlier than that of the idler.

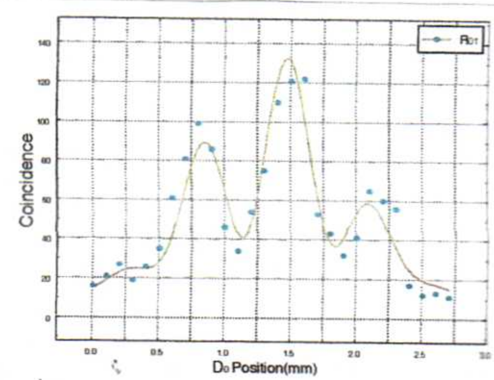


FIG. 3. R_{01} ("joint detection" rate between detectors D_0 and D_1) against the x coordinates of detector D_0. A standard Young's double-slit interference pattern is observed.

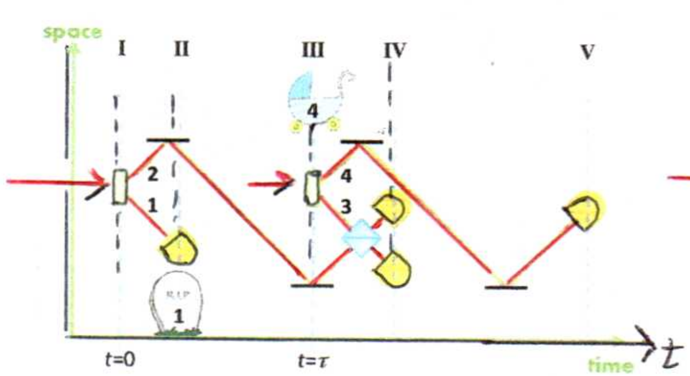


FIG. 1. (color online). Time line diagram. (I) birth of photons 1 and 2. (II) detection of photon 1. (III) birth of photons 3 and 4. (IV) Bell projection of photons 2 and 3. (V) detection of photon 4.

Entanglement Swapping [Meg] 2012

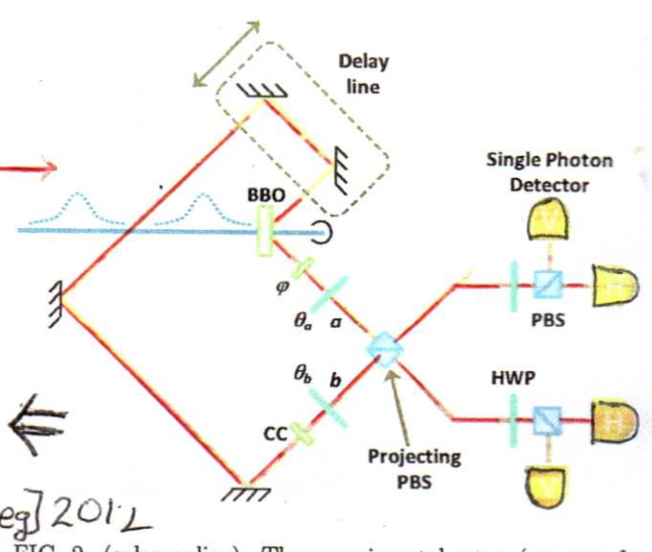


FIG. 2. (color online). The experimental setup (see text for details).

Appearances of Retrocausality in Entanglement Experiments

Dave Peterson 11/2/18 –11/15/18

Four early classic journal articles on photon entanglement are sampled for discussion. All use "SPDC" crystal down-conversion to create entangled photon pairs from UV laser beams. A focus is on "quantum erasures" of double slit interference patterns with and without significant delayed choice. Delayed decisions in one system can determine whether interference occurs or not in another system. Counter to intuition, event times, spatial separation and order of events have no effect on outcome – as if some sort of "backward-in-time" quantum communication could occur between entangled particles. One sampled article shows that when identical photons travel on identical paths, double laser beam passes through a crystal may no longer be independent; and four created proto-photons emerge as just two. The efficiency of crystal pair production can also be enhanced or reduced by external optical interferences. These articles are from 1994 to 2001, but a later selection from 2012 is also included: "Entanglement Between Photons that have never Coexisted" (entanglement swapping). Entanglement correlations can exist with temporal separations.

{See **Appendix** for discussions for and against some sort of "back-in-time" quantum information transfer}

Question: can one learn the new quantum mechanics of entangled particles by reading published journal articles? They are very concise and intended for specialists in increasingly separated sub-fields. Every word counts. Non-specialists (almost all of us) then have to struggle to decipher these publications and supplement them with some related literature. Perhaps a good up-to-date textbook would be a better approach. Is there one?

Many entangled particle experiments reveal distant correlations in space-time that seem to require "faster than light communications" in violation of special relativity (spooky action at a distance). It has occasionally been suggested that the quantum entanglement correlations may be due to back-and-forth communications in time that emulate instantaneous communication between measuring systems. The "no-signaling" theorem of quantum mechanics firmly says that no classical-information can be transmitted backwards in time. A possible solution is to suggest that "quantum information" operates in a different sub-realm and transmission can be bi-directional in time. This has been called "retro-causal," but the term causality refers to cause-and-effect relations in the classical world. There are now a vast number of entanglement experiments that disturb traditional thinking and give the "appearance" of some sort of retrocausal communication. This note discusses a few key examples.

Prior to the 1960's, people knew that quantum mechanics was weird, but they weren't overly concerned about the specter of non-locality. From the 1960's through the 1980's there was a "second quantum revolution." This was initiated by John Bell with his articles about two entangled particles; and that was then followed by a huge and accelerating cascade of studies exploring the new frontier of entangled particles. The goal of this paper is to mention some of its new discoveries and focus on a sampling of just a few entanglement experiments of this frontier.

The short list of new topics introduced in the last half of the 20th century include: Bell Inequalities, Mach-Zehnder interferometry, Quantum Erasures, Delayed Choice

Quantum Erasure, Hanbury Brown & Twiss effects, Quantum computing, Quantum cryptography, Quantum Teleportation, Entanglement Swapping, Coherent states, GHZ states, Hong-Ou-Mandel effects (HOM), Bohmian Mechanics, “Entanglement and nonlocality in multi-photon systems,” effective de Broglie wavelengths of n particles, photon orbital angular momentum (OAM), Trapped Ions, Decoherence, and other numerous topics in Quantum Optics. There are many experiments incorporating some of these topics that demonstrate clear non-local behaviors.

As a prequel, consider an old 1927-type puzzle for a simplest one-photon example: place cool radioactive nuclei at an origin and observe gamma rays heading toward a 180 degree curved detection screen. For each decay, there is only one spot on the screen; and all the other “possible” spots are then denied actuality. There is also a reaction kick backwards on the nucleus. Are the directions of the back-kick and forward spot location correlated? – yes, of course, due to conservation of momentum. But one can only see the correlation after the screen spot has appeared some time later after the recoil. When does the kickback occur? -- immediately at time zero before the spot shows on the screen. It is a challenge to imagine how that can happen.

Most of the following discussions require looking at the actual journal articles for **picture Figures** and details (available on Web as listed below).

A: The Essence of the Herzog Experiment -1994 Ref: [**Frust**].

[The Layout geometry of the test affects the efficiency of down conversion!].

See Figures in: T.J. Herzog, J.G. Rarity, H. Weinfurter, and A. **Zeilinger**, “Frustrated **Two-Photon** Creation via Interference,” PRL 72, #5 31 Jan 1994 OR https://www.univie.ac.at/qfp/publications3/pdffiles/1994-04.pdf

Given: A UV laser beam passes through a down-conversion crystal from the left, reflects from a mirror back through the crystal again allowing the possibility of two pairs of lower energy photons. {Topic of e-mail group discussion for Boulder Library Cosmology group, Sept. 2018}

Usual Intent: Most of these experiments with forward plus reverse throuputs wish to generate **four** “real” photons on four separate outgoing paths: two to the left and two to the right. And this would happen with very low efficiency [probability squared or amplitude $|\alpha|^4 \sim 10^{-13}$ per UV-photon input]. But, in this case, the pair productions are not independent.

Herzog’94: The right side photons are separately reflected backwards --folding the four paths into only two precisely distinct paths to the left and eventually producing two “real” photons at the final detectors. Immediately to the left of the crystal, consider the special simple case of created photons overlapping on the same paths at the same time and same distance with all differential phases set to zero. The two pairs become entangled and their amplitudes can now interfere. Because they are bosons, they interfere with a + sign: $\psi = \alpha\, [s_1 i_1 + s_2 i_2] \rightarrow s_1 s_2$ & $i_1 i_2 \rightarrow$ “s & i” with path identity before D_s and D_i detectors {ψ is symmetric under inter-changing of pairs}. Then probability goes as $\langle\psi|\psi\rangle = 4\alpha^2$ enhanced above just α^2 because of boson entanglement and how that affects quantum down-conversion events.

IF the final path photons had instead been fermions, then interference would have been anti-symmetric under exchange: $\psi \propto [\,(s \cdot i) - e^{i\phi}\,(s \cdot i)\,]$ so that $\psi^*\psi \propto\ 2\text{-}2\cos\phi = 0$ for phase $\phi = 0$ (identical particle overlap would have been Pauli-excluded!). But then, of course, we wouldn't have been able to generate two photons from one in the first place {lepton number is conserved}.

Brainteaser: Detection, D, collapses the superpositions and randomly says that only one of the creation pairs "actually occurred." The measured enhancement of detected photon rates was due to the possibility of either of the creation pair events occurring. Detection happened more than 10^5 wavelengths after the crystal pair creation event #2, and the first pair-creation event #1 happened ~300,000 wavelengths prior to that [$t_D > t_2 > t_1$]. These three events mutually coordinate without caring about their order in time. It would appear that the quantum world of possibilities possesses some "reality" below classical reality.

{Some Details: UV pump laser beam λ_p =351.1 nm passes through an $LiIO_3$ non-linear crystal with "phase matching" inside the crystal → red λ_s= 632 nm "signal" photons exiting the crystal at a shallower angle than companion 789 nm "idler" photons. The final trajectory to a signal detector, D_s, is configured to only allow signal photons. There is fine stepper motor control of reflecting mirrors for phase interference scanning output. Coincidence counts of both D_i and D_s detectors together reveal the original 351 nm pump waves.}

B: **"A Delayed Choice Quantum Eraser" [Kim, 1999]:**

See figures in: Wikipedia /wiki/Delayed-choice_quantum_eraser (Kim 1999). Original article, "A **Delayed choice quantum eraser**," PRL 1999-2000 pdf at https://arxiv.org/pdf/quant-ph/9903047.pdf Yoon-Ho Kim, et. al.,

Despite the complex experimental figure, the physics here is just standard Young's phase interference -- but coupled with entangled partners. Two ray interference at a transverse-moving detector can be destroyed by removing a beam splitter in a distant complimentary system.

"The which-path or both-path information of a quantum can be erased or marked by its entangled twin even after the registration of the quantum." Nothing in these tests is really being "erased"—they simply defy our beliefs that later events should not be able to affect earlier events. For entangled particles, time-ordering is often violated – our beliefs were not valid. Traditional wording of "which-path" knowledge preventing interference is "physically unsatisfying and needs some elaboration.

In this experiment, a pump laser excites two sets of pair creations at the same time with one pair traveling through a close-up double slit hole and the other through the other slit (slits A and B). One ray from each slit travels to a relatively nearby single detector D_o which can be moved sideways to show possible interference patterns from gathered statistics of clicks. The other entangled ray from each slit goes into a prism and then through beam splitters and mirrors into four more remote detectors such that outer detectors (D_3, D_4) see unique paths identifying "which slit" origins (A and B, and hence no possibility of interferences). These detectors are 2.5 meters or 8 ns farther out than Do. The other two paths simultaneously go through a common beam splitter into detectors D_1, D_2. These interfere and then allow traditional interference in the entangled paths to the D_o detection. The "both-path" (1,2) interference erases "which-path" (3,4 =A and B) marking each s and its "i".

The tests output four "joint detection rates" R_{01}, R_{02}, R_{03}, R_{04} between D_o and the other four detectors. The joint detection amplitudes for the outer detectors only have one contributing amplitude, but the others have interfering amplitudes A(1&A) + A(1&B) and A(2&A)- A(2&B).

The "Schematic of the experimental setup" [Figure 2 in the article] appears complicated. But if one thinks of photon paths as flexible wires that can be bent, then Figure 2 can be converted back to a simpler conceptual picture in Figure 1 (that is important for understanding). It is also very important to see that the paths from detector D_1 to BS_c beam splitter and to D_2 have double lines (both red and blue in the color drawing shown in the Wikipedia version of the Kim article). The experiment shows four-photons as double sets of down-conversion pairs of photons passing immediately through a double slit that could produce usual screen interference wave profiles [ensemble of detector D_o clicks versus translation distance x] while their complement photons go into a distant interferometer with three beam splitters going to four detectors. What happens at a later time in this interferometer affects whether early time screen interference is seen or not.

Suppose we consider quantum-wave light paths to be like bi-directional transmission cables that can transport quantum information forwards and also backwards in time (not like classical cables). A new-perspective might be that when screen-interference does occur, the diamond shape of the Mach-Zehnder interferometer path segments forms a functioning "closed circuit." But if just single path outer detectors click (photon dump), then no interference occurs. One can then suspect that the term "which path information" goes with "no closed path circuit." A later photon dump makes its entangled photon simply dump too – and interference can occur when late time photon clicks have a closed circuit connection to the D_o screen-interference clicks.

This report produced a lot of controversy about what's "really" going on {e.g., [Bram], [Fank] }.

C: An early 2002 "simple" Case: Two photon quantum eraser:

{This involves only two photons but is not quite simple since it interchanges three different basis sets for polarizations: Linear, Diagonal, and Circular. So extra math is needed}.

In this test, two-slit interference is activated. Then it is destroyed and then activated again thus erasing previous non-interference.

See **Figures** in: S. P. Walborn, et. al., **"A double-slit quantum eraser",** Phys. Rev. A. 65 (3):033818 https://arxiv.org/pdf/quant-ph/0106078.pdf , 2001. Also chosen as introduction in Wikipedia, https://en.wikipedia.org/wiki/Quantum_eraser_experiment

[Experimental Details: Two entangled photons are created using Type II nonlinear birefringent crystal { β BaB_2O_4 or "BBO" } spontaneous parametric down conversion "SPDC" that outputs an ordinary and an extraordinary photon o-ray and e-ray. One ray will have vertical polarization and the other horizontal polarization, V or H (depending on setup). Output light exits in two tilted cones for H and for V that overlap at two crossover points so that we cannot differentiate between H or V. The pump laser input has UV-wavelength 351 nm → γ + γ at 702 nm exiting at 3^o. A lower photon path or "**signal**" s ray (label "L" or "s") travels 42 cm to a double slit having opening and spacing both at 200 microns. Upper and lower downstream detectors are located about a meter away. There are coincidence counters for D_{up} and D_{lower} (or D_s) which can move laterally to show interference fringes from the double slit. But note that correlation can't be stated until after there is data available to correlate.

Initially, in the **Walborn** journal article, the created photons are linearly polarized in a Bell state of entanglement:

$|\Psi^+\rangle = |\psi\rangle = (|H\rangle_s|V\rangle_u + |V\rangle_s|H\rangle_u)/\sqrt{2} \equiv \kappa\,(HV + VH)$, in simpler form, ordered first by lower signal "s"- beam and then by the upper beam to detector: product order $|L\rangle|U\rangle$ **or s u,** with U upper path lacking any interferometry in all cases. We let $\boldsymbol{\kappa} = 1/\sqrt{2} \sim 0.707$ for convenience of notation. This is a natural starting point for type II SPDC.

The lower s-beam illuminating the double slit gets divided into two exit beams s_1 and s_2—for slits 1 and 2 . These are each still entangled with the upper beam U (or u). $\Psi_1 = \kappa(H_{s1}V_u + V_{s1}H_u)$ and also $\psi_2 = \kappa\,(H_{s2}V_u+V_{s2}H_u)$. The usual two-ray interference can result as probability $P_s \propto (1+\cos\Delta)$ where Δ is the phase difference between the paths s_1 and s_2. The upper beam goes to a movable detector that could reveal an interference pattern in its count statistics versus lateral displacement.

Next, we position quarter-wave-plate (QWPs , λ/4) circular polarizers in front of one of the lower slits for "right handed or counter-clockwise" CCW polarization and the other CW for left- ℓ handedness. This "marking" of the slits by r and ℓ polarized ray labels causes the interference pattern to vanish due to new "which path" information.

A later modification is to finally add a 45 degree polarizer in place before the other upper path detector, D_u. Since "diagonal" polarization d = κ (H+V), this allows both H and V light (or r and ℓ light) to pass through both slits and allows interference again.

> "*Entanglement ensures a complementary diagonal polarization in its partner, which passes through the double-slit mask. This alters the effect of the circular polarizers: each will produce a mix of clockwise and counter-clockwise polarized light. Thus the second detector can no longer determine which path was taken, and the interference fringes are restored."*

<u>Now some math</u> (not shown in Wikipedia—and elaborating Walborn assumptions). This is an aside exercise in **relating <u>three types</u> of polarization basis sets**. All are used in this journal article.

<u>Right circular polarization</u> is defined as **$|r\rangle = r = \kappa\,(H+iV)$, and $\ell = \kappa\,(H-iV)$.** The addition of H and V out of phase produces a "cork-screw" or helical E field.
For <u>diagonal</u> polarization we can switch to a new diagonal basis: d and δ for $+\pi/4$ tilt and $-\pi/4$ tilt transverse polarization. These are orthogonal, so clearly $\langle d|\delta\rangle = 0$ as also do $\langle H|V\rangle$ and basis $\langle r|\ell\rangle = 0$. At will, we can translate between these three choices of polarization bases.

Solving for H and V from above we get: in circular polarization basis, $H = \kappa\,(\ell+r)$ and $V = \kappa\,(\ell - r)$.

The original polarization states, H and V, have no angular momentum, while circular polarization does. Conservation of spin requires that ℓ and r terms be balanced for zero net spin.

Wave functions have phase α in $e^{i\alpha}$ as a gauge degree of freedom, -- meaning that it isn't really functional. For diagonal α =45 degree rotation, we consider $e^{i\pi/4} = \cos(\pi/4)+i\sin(\pi/4) = 0.707 + i\,0.707 = \kappa + i\kappa = \kappa(1+i)$. A term $\kappa\,(1-i) = -i\kappa\, e^{i\pi/4}$ is also used in what follows below. {The wave function set is still "essentially' the same with or without the phase}.

The definition for **<u>diagonal d</u>**$=\kappa\,(H+V)$ and $\delta = \kappa\,(H-V)$.

{Walborn somehow seems to reverse these symbols}

Or, $H = \kappa\,(d+\delta)$ and $V = \kappa\,(d-\delta)$. Using these three sets of relations {for H,V,r,ℓ ,d,δ) yields:

$|r\rangle = r = (\tfrac{1}{2})(d(1+i)+\delta(1-i))$. Similarly, $|\ell\rangle = \ell = (\tfrac{1}{2})(d(1-i)+\delta(1+i))$,

As a quick check, $\langle r|r\rangle = (\tfrac{1}{2})(d^*d+\delta^*\delta + (i-1)\,|d\delta\,|\,) = (\tfrac{1}{2})(1+1+0) = 1$ = 100% probability.

Solving for the diagonal basis in terms of the circular polarization basis gives:
$d = (\tfrac{1}{2})(r(1-i) - \ell(1+i)$ and $\delta = (\tfrac{1}{2})(r(i+1)+\ell\,(1-i))$.

To understand and simplify these somewhat strange and complex forms, we now consider and remove phase, α. Factoring out the phase, we have $\delta = \kappa e^{i\alpha}(r - i\ell)$ and $d = \kappa\, e^{i\alpha}\,(\ell - ir)$.

But, the overall phase is not functional and can be ignored and deleted:
so $d \cong \kappa\,(\ell - i\,r)$ and $\delta \cong \kappa\,(r - i\ell)$.
Then, $\mathbf{r} = \kappa\,(d - i\delta)$ and $\boldsymbol{\ell} = \kappa\,(\delta - id) = -i\,(d+i\delta)$.

<u>[Caution</u>: If you read Walborn **<u>and</u>** its shortened Wikipedia version together, note that Wikipedia simplifies and <u>alters</u> the Walborn scenario: WIK <u>begins</u> with UL order ($\ell r + r\ell$) for the lower quantum state with its resulting circular polarizations. Then for U, filtering out ℓ leaves term $r_U\ell_L$, If U measures a linear H, then $kH(\ell + r) + ikV(r-\ell)$ leaves $(\ell + r)$ lower so slit rays can interfere.
But Here we follow Walborn who **<u>begins</u> with the Bell $\psi = |\Psi^+\rangle = \kappa(HV+VH)$,** a different starting point that doesn't quite agree with the assumptions in Wikipedia. Order here is LU instead of UL. The L slits split ψ_1 and ψ_2 which are considered separately. Then a DIAGONAL filter (not just H) is applied to U leaving diagonal d interferences for s_1 and s_2.

As example, another original Bell state $\psi = \kappa$ (HH – V V) = (½)[(ℓ+r)(ℓ +r) – {(ℓ-r)(ℓ-r)} = (r ℓ + ℓ r) - {like the Wikipedia case. **But** Walborn claims the state $\psi = \kappa$ (HV+VH) instead.]

Now, … finally! .... we can return to the quantum erasure experiment now with a diagonal polarizer inserted in the upper path. What does it do and how does it affect the lower path?

[UL ordering] In the upper path we insert a **"diagonal" filter** – H tipped up by 45 degrees. Then, for the Upper path, we switch to diagonal representation. ψ = (½)[(d-iδ)ℓ + (δ –id)r] with the lower path set up to have circular polarizations, ℓ and r.
Selecting polarizer d in U forces a collapse to (½)d (1ℓ -i r) which is a superposition of both left and right rays – and that can go through the circularly polarized lower slits!
So, interference has been turned back on.

In more detail, **the two rays** from the slits 1 and 2 of the double slit mask have separate wave functions: $\psi = \kappa (H_s V_u + V_s H_u) \rightarrow \psi_1+\psi_2$: ψ_1 with $s \rightarrow s_1$, and ψ_2 with s_2.

Then with the two different quarter wave plates in front of the slits, we convert s to circular basis notation. [LU ordering]
Rays: $\psi_1 = k(\ell_{s1} V_u + i r_{s1} H_u)$, and $\psi_2 = \kappa (r_{s2} V_u - i\, \ell_{s2} H_u)$.

This choice ensures that the orthogonal form $\langle\psi_1|\psi_2\rangle = 0$ and $\langle\psi_1|\psi_1\rangle = 1$.

Now switch from H&V to d and δ, where $H = \kappa(d+\delta)$, $V = k(d-\delta)$,

Also $r = \kappa (d-i\delta)$ and $\ell = \kappa (\delta - id)$.

Expand Rays. Add and Reorganize

$\psi = \kappa [(½)(\delta_1 + i\delta_1) - i\delta_2]\, d + (d_2 + id_1 - i\delta_1)\, \delta$

Then the U d-polarizer on U path collapses the d term and selects

$\psi = \kappa [(½)(\delta_1 + i\delta_1) - i\delta_2]$ for L – a combination of the two rays together.

So, we now have clear two ray interference in the lower path.
We have erased the non-interference!!

{Walborn doesn't delete the $e^{i\alpha}$ phase and leaves the (1-i)/2 's intact.
And his final equation #14 has: $|\psi\rangle =(½)[(d_{s1}-id_{s2})\, d_u + i(\delta_{s1}+i\, \delta_{s2})\delta_u]$
{I see no way to obtain this equation—so we have an important disconnect!}
If U filter has d or δ , **the two lower s rays interfere either way**.
Both math solutions re-activate interferences.

D. Entanglement Between Photons that have Never Coexisted.

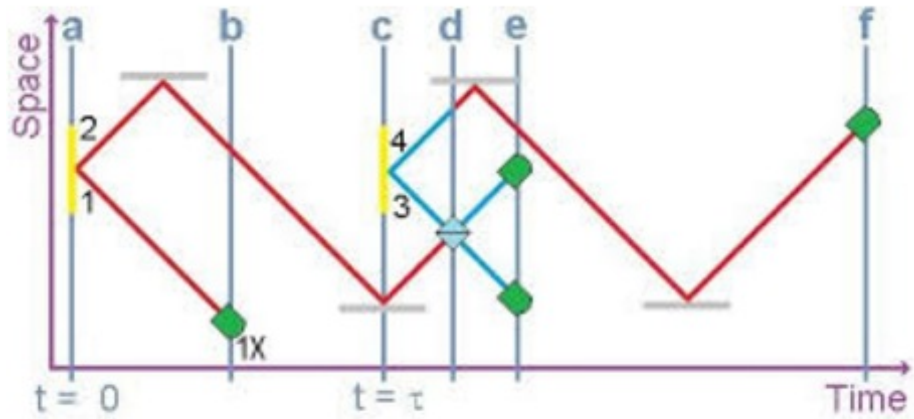


Figure: **"Experimental Entanglement Swapping: Entangling Photons That never Co-existed"** [my favorite entanglement picture].

A UV laser beam enters from the left through an SPDC crystal (yellow) where it can output two lower energy entangled photons. Photon 1 goes straight to a detector (green) while 2 goes to a Polarizing Beam Splitter (PBS). The laser beam passes straight through the first crystal into a second one where two new entangled photons are created. Photon 3 and 2 coincide exactly at time " **d"** on the PBS with 2 and 3 becoming entangled and 2 then getting detected. Then 4 continues to its final detection. **1 and 4 are now entangled** although 4 originated well after the creation and detection of photon 1. {The transactional interpretation of QM would say that the detection of photon 4 sends a confirmation wave back in time from photon 4 at f along a convoluted path to the sources of the offer wave thus establishing correlation. But in this case there are two sources.}

The actual experimental setup was somewhat different and more complex than the basics of the figure above. Rather than two crystals, there was only one with two laser pulses passing through it separated by $\Delta t = \tau = 105$ ns. The first surviving photon 2 (at initial time t=0) goes through a long delay line (31.6 meters) so that it can have an encounter event with one of the second photon pairs #3 (at $t = \tau$). Photon γ_o and γ_τ (2 and 3) are then combined at the PBS and projected onto a Bell state. In path 1 is a polarization rotator (HWP at angle θ_a) and another along path 2 (angle θ_b). Setting these to zero allows the usual polarization states: $|hh\rangle$, $|hv\rangle$, $|vh\rangle$ and $|v\,v\rangle$.
The single photon polarization detectors at b, e, and f help identify these states.

Now we need to address Bell States and SPDC (spontaneous parametric down conversion of a laser pump photon into two lower energy entangled photons). The four Bell states are: $|\Psi^\pm\rangle = \kappa(|H_aV_b\rangle \pm |V_aH_b\rangle)$ and $|\Phi^\pm\rangle = \kappa(|H_aH_b\rangle \pm |V_aV_b\rangle)$ where $\kappa = 1/\sqrt{2} \sim 0.707$ (for convenience of expression).

The most commonly discussed Bell state is $|\Psi^+\rangle$ produced by type-II SPDC with exact phase matching. In general the state can be $\psi = (|HV\rangle + e^{i\phi}|VH\rangle)/\sqrt{2}$ [Brida]. For mismatching values of $\pm\pi/L$ (length of the crystal), the output pair is $\psi = |\Psi^-\rangle$. Similarly, a type-I SPDC BBO crystal can produce a continuous set of polarization-entangled states from $|\Phi^+\rangle$ [for $\theta = 0$] to $|\Phi^-\rangle$: $\psi = (|HH\rangle + e^{i\theta}|V\,V\rangle)/\sqrt{2}$. PDC crystals are birefringent with polarized rays having two different indices of refraction. For type II, an ordinary output ray is V and is called "idler," and the extra-ordinary ray is H and called a "signal."

In this paper, tilting of a compensating crystal in the downwards paths 1,3 "control the phase φ of the state, e.g., for $\varphi = \pi$ the resulting state is $|\Psi^-\rangle$. The four photon state is $|\Psi^-\rangle_o \otimes |\Psi^-\rangle_\tau$ at times zero and tau. Including the staggered delay times for upward path photons 2 and 4 modifies this to four terms summarized by time 0, τ ,τ, and 2τ : $|\Psi\rangle 4 = |\Psi^-\rangle_{o,\tau} \otimes |\Psi^-\rangle_{\tau,2\tau}$ which ends up containing ± terms over all four Bell states.

The single photons 2 and 3 combine at the PBS, ~(h±v)⊗ (h± v) → terms like hh, vh, hv, and vv. They arrive as photon-bosons at the PBS at exactly the same time so forming hh or vv terms both either transmitted or reflected (like HOM effect). So these get projected onto a $|\Phi^+\rangle_{\tau\tau}$ or a $|\Phi^-\rangle_{\tau\tau}$ state.

Then characterizing the first and last now entangled photons 1 and 4 requires gathering data for a quantum state tomography (QST). This is "the process of reconstructing the quantum state (density matrix) for a source of quantum systems by measurements on the systems" [Wikipedia]. The results of this agree with the states $|\Phi^+\rangle$ or $|\Phi^-\rangle$ when plotted as a 4x4 square of {hh, hv, vh, vv} x {hh, hv, vh, vv}.

References:

[Frust] T.J. Herzog, J.G. Rarity, H. Weinfurter, and A. **Zeilinger**, "Frustrated Two-Photon Creation via Interference," PRL 72, #5 31 Jan 1994. Article in: https://www.univie.ac.at/qfp/publications3/pdffiles/1994-04.pdf And more discussion found in:

[Herzog]: Thomas J. Herzog, Paul G. Kwiat, Harald Weinfurter, and Anton Zeilinger, "Complementarity and the Quantum Eraser," Physical Review Letters, Vol. 75, No. 17, 23 October, 1995, Institut fur Experimentalphysik, Universitat Innsbruck Paper Also at: http://isis.ku.dk/kurser/blob.aspx?feltid=17775 (ISIS, Copenhagen University). This includes [Frust].

[KIM] Wikipedia /wiki/Delayed-choice_quantum_eraser (Kim 1999). Original article at https://arxiv.org/pdf/quant-ph/9903047.pdf Yoon-Ho Kim, et. al., A **Delayed choice quantum eraser**, PRL 1999-2000

[Walb] S. P. Walborn, et. al., **"A double-slit quantum eraser",** Phys. Rev. A. 65 (3):033818 https://arxiv.org/pdf/quant-ph/0106078.pdf , 2001. Also chosen as introduction in Wikipedia, https://en.wikipedia.org/wiki/Quantum_eraser_experiment

[Meg] E. Megidish et al., "Entanglement Between Photons that have Never Coexisted," https://arxiv.org/abs/1209.4191, 2012 – Swapping.

[Brida] G. Brida, et al., "Generation of different Bell states within the SPDC phase-matching bandwidth, arXiv:0706.0848, 2007

Other related journal articles include:

[Herbst] . Thomas Herbst, et. Al., Anton Zeilinger, "**Teleportation of entanglement over 143 km**," https://arxiv.org/abs/1403.0009 PNAS 112 no 46 Nov. 2015 (or Letter Published: 05 September 2012).

[Gaas] (Gaasbeck, 2010) https://arxiv.org/pdf/1007.3977.pdf Demystifying the delayed choice experiments
http://strangepaths.com/the-quantum-eraser-experiment/2007/03/20/en/, The quantum eraser experiment xantox, 20 March 2007 in Physics

[Zeil] Anton Zeilinger, **"Entanglement by Path Identity"** (arXiv1610.00642).

[Feyn] Feynman Lectures in Physics, 1965: {elementary introduction to Bose particles in Lectures III p 4-4}:

[Anil]: Anil Ananthaswamy, **Through Two Doors at Once**, The elegant experiment that captures the enigma of our quantum reality, 2018

[Ma] Xiao-song Ma, et. al., Quantum erasure with causally disconnected choice, arXiv:1206.6578, 2013

[Walther] Philip Walther, Entanglement and Nonlocality in Multi-Photon Systems, 2005

[Pan]: Philip Walther, Jian-Wei Pan, et al., **De Broglie wavelength of a non-local four-photon state,** Nature 429 13 May 2004 p158.

[Ou]: Ou, Zou, Wang, Mandel, "Observation of nonlocal interference in separated photon channels, PRL 63, July 16, 1990

[Bram] Bram Gaasbeek, "Demystifying the Delayed Choice Experiments," arXiv:1007.3977 2010.

[Fank] Johannes Fankhauser, "Taming the Delayed Choice Quantum Earser," arXiv:1707.07884 2018.

"Lab Tools for Quantum Physics" : https://www.qutools.com/newsletter/ and "Smart Tools for Quantum Optics":

https://mail.google.com/mail/u/0/#inbox/FMfcgxvzLWzdgVthPCFJWSVBzcXRMgjv?projector=1&messagePartId=0.1

Charles H. Bennett and others, "Quantum Nonlocality without Entanglement," https://arxiv.org/abs/quant-ph/9804053

Notes:

Define Entanglement:

"An entangled system is defined to be one whose quantum state cannot be factored as a product of states of its local constituents; that is to say, they are not individual particles but are an inseparable whole." For two entangled photons, the polarization has to be indeterminate -- not horizontal, not vertical, not circular, just a blank that has yet to be filled in. Entanglement represents property conservation by superpositions of alternate possibilities. For bipartite systems, entanglement implies non-locality – nonlocal correlations where one constituent cannot be fully described without considering the other.

Entanglement Properties can include polarizations: H, V, $|+\rangle$, $|-\rangle$; electron spin up or down, path entanglements, and boson occupation numbers like the NOON states: $|2,0\rangle + |0,2\rangle$**.** This occurs in the "HOM" effect when photon encounter a beam splitter and arrive within the coherence time (they "merge"). There is also "entanglement by path identity." (Zeilinger). It is rarely mentioned, but **entanglement superpositions are a way to enforce conservation laws given a world of possibilities.** The possibilities here are that A could see an H or V but the SPDT output H and V relationships are at least preserved in the outcome.

Discussion of possible "retrocausality" in quantum mechanics:

PRO: Perhaps the earliest suggestion of a back-in-time explanation for entanglement correlations was by Costa de Beauregard in 1953. This is called the "Parisian Zig-Zag." Since 1986, John Cramer has been the primary source for "The Transactional Interpretation of Quantum Mechanics" (TIQM) based on Wheeler-Feynman absorber theory of electromagnetism. This is an important arena with thirty years of literature; so I will leave this subject for, "Look on Google." People who don't study this will have limited imaginations. They might imagine that a quantum state between emitter and absorber is a single time directed thing while TIQM says it is a back-and-forth "process" of transactions ending in a "hand-shaking" agreement to transfer a quantum.

Similarly, Yakir Aharonov's theory of weak measurement within the Two-State-Vector Formalism (TSVF) "is based on the assumption that quantum interaction involves a combination of past and future state vectors and also involves back-in-time communication" (again, see Google for details). Huw Price says that if the quantum world is time symmetric or reversible land is also real, then retrocausal influences are required. (2012, Phys.org.news 2017). Mathew Pusey and Matthew Leifer agree. If QM is incomplete, retrocausality might complete it.

Then there is Roger Penrose ("Road to Reality" p.603) who proposes that quantum information can flow Zig-Zags both forwards and backwards in time. Entanglements are created locally (at the same place, same time, from same atom, from same input photon, necessarily close together). There is a distinction between "ordinary or classical information" (evolving only forwards in time) and quantum information which is different and referred by him as "Quanglement" – "it is appropriate that the name suggests entanglements" and it can flow both forwards and backwards in time (like zig-zags). A "guiding principle behind Penrose's twistor theory is quantum non-locality." [p963]; but, despite much study, it is still in a primitive state of development.

"Most if not all interpretations of QM involve action outside the light cone: Bohm uses pilot waves; Copenhagen has wavefunction collapse; TIQM has Wheeler-Feynman absorber theory, which uses a different kind of pilot wave; Consistent Histories assigns probabilities to events in the past; Many Worlds has interaction between alternative universes." (Phys.org.news). "If retrocausality is allowed, then the famous Bell tests can be interpreted as evidence for retrocausality." (phys.org)

CON:
"With the relational interpretation you need neither action at a distance nor retrocausality.. just give up the metaphysics of ontological realism. Or, one may merely assume that multi-particle physics takes place in "configuration space" (e.g., a particle moving along the x axis and another along the y axis corresponds to one point on an x,y,t space. [But why is that allowed? They might be jointly connected by a "zig-zag" path]. Another choice is to simply deny any kind of quantum realism – that is consistent with the Copenhagen dogma that there is no reality below measurement. Two articles on the quantum eraser both deny the need for retrocausality [Fank][Gass]. It is easy to deny a sub-quantum reality simply by saying that psi is a probability amplitude—how real can that be? Well, one could think of psi as some sort of sub-real matter wave

analogous to the electric field amplitude so that when it is squared it is like energy density (psi*psi) -- possible real existence density. In general we get some possibility intensity. **Then step two** after that is to add a "Principle of Random selection" to definite outcomes ("PRS", mechanism unknown, but equivalent to a transaction hand-shaking agreement). Then we get a probability **as if** the wave were a probability amplitude (for all **practical purposes, FAPP).** The PRS selects a target and cancels the other possibilities. Because of PRS pure randomness, there is never a classical cause and effect and there is "no signaling."

THE DENSITY MATRIX

DAVE PETERSON

ABSTRACT. Quantum mechanics can be formulated either by a density matrix formalism or by the more common state vectors belonging to a Hilbert space. The density matrix is increasingly finding more relevance and application. For example, an entangled state can be "pure" (perfect correlation between two systems) while each of its individual systems sees "mixed states" (such as unpolarized light). This can be discussed by reducing a density matrix from the combination into density matrices for each part separately.

1. Introduction

The density matrix concept was introduced separately by Lev Landau and John von Neumann in 1927 to describe statistical ensembles of systems. It has special use in problems with entangled systems and in discussions of decoherence and quantum entropy. It can even be considered as an "interpretation" of quantum mechanics: Steven Weinberg [1] recently proposed that we rely on the density matrix as the description of reality instead of physical states in terms of ensembles of state vectors. The density matrix has the advantage of applying not just to the usual "pure states" of most introductory texts on quantum mechanics but also to mixed states given by probabilities and not just quantum superpositions of pure states. An example of a pure state is vertically polarized light, $|V\rangle = \frac{1}{2}(|R\rangle + |L\rangle)$, in-phase superposition of right and left circularly polarized light. In contrast, Unpolarized light is a mixed state statistical ensemble with 50% probability of being R or L or also polarizations horizontal or vertical.

Since there are two base states here, the density matrix, ρ, would be represented by the simplest case of 2×2 matrices with a general form [2]:

$$\rho = \begin{pmatrix} a_{11} & a_{12} + ib_{12} \\ a_{12} - ib_{12} & a_{22} \end{pmatrix} \rightarrow \begin{pmatrix} A & B + iC \\ B - iC & 1 - A \end{pmatrix}, \ e.g., \ \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{1}$$

The density matrix in general has the following requirements:

1) $\rho^\dagger = \rho$, the density matrix is Hermitian (equals the complex conjugate of the transpose about the diagonal). This means that the diagonal elements are real, and the off-diagonal elements are complex conjugates.
2) $Tr\,\rho = 1 = 100\%$ ('Trace' is sum of diagonal terms), so if $A = a_{11}$ then $a_{22} = 1 - a_{11}$

Date: May 15, 2014.
email: davepeterson137@gmail.com. Paper updated to May 22, 2014.

$= 1 - A$.
3) All eigenvalues λ_k of ρ must be nonnegative, $0 \leq \lambda_k \leq 0$ or $\rho \leq 1$.
4) For a pure state, $\rho^2 = \rho$, so $Tr(\rho^2) = 1$, but a mixed state has $Tr(\rho^2) < 1$.
5). Expectation values for an operator A can be calculated using $\langle A \rangle = Tr(\rho A)$ [4].
6). The density operator evolves in time as: $i\hbar \frac{\partial \rho}{\partial t} = [H, \rho] = H\rho - \rho H$.

Eigenvalues of a matrix, λ_k, are found as usual by solving the "Characteristic Equation," polynomial $det|\rho - \lambda I| = 0$ (subtract lambda from diagonal terms). For the density matrix form above, this gives $\lambda^2 - \lambda + A - A^2 - B^2 - C^2 = 0$. Solving by the quadratic formula and having $\lambda \geq 0$ requires that: $(A - \frac{1}{2})^2 + B^2 + C^2 \leq \frac{1}{4}$ [1]. This can be plotted as a unit Ball B^3 in Figure 1 with center at $A = 0.5$ and radius 0.5. The boundary of the ball (or 3-disk) is the two-sphere of pure states [2], and the interior is mixed states. An arbitrary pure state is defined by the latitude and longitude on the sphere. For bases like up and down, u and d, we can have: $|\psi\rangle = \sqrt{A}|\uparrow\rangle + \sqrt{1-A}e^{i\phi}|\downarrow\rangle$ (where ϕ here is the polar angle). Photons that pass through a vertical polarizer would have $A = 1$ with all other terms zero; ie., only the pure state at the north pole of the ball with $a_{11} = 1$ in equation (1).

Instead of Dirac "inner-product" order "bra-ket", the density operator (matrix) is defined in terms of "outer products" like "ket-bra" $|\psi\rangle\langle\psi|$ [3]. Suppose we have a quantum state that isn't known well. But there is some probability, p, that the state might be $|\psi\rangle$ and some probability, q, that it might be $|\phi\rangle$. Then the density matrix is defined as
$\rho = p|\psi\rangle\langle\psi| + q|\phi\rangle\langle\phi|$ [3]. If both p and q (etc.) are non-zero, we have a mixed state; but, If only one of these terms is given (say $p = 1$, $q = 0$), then we have a pure state, $\rho = (1)|\psi\rangle\langle\psi|$. Notice that for a pure state, $\rho^2 = |\psi\rangle\langle\psi| \cdot |\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \rho$. [4] Geometrically, if we plot the pure state points $|\psi\rangle$, $|\phi\rangle$ on the Bloch sphere of figure 1, then the location of the mixed state given by ρ is a point along the chord joining the two outer points at relative distances given by the probabilities, p and q. That is, ρ will lie somewhere inside the sphere. The collection of all such points is the solid ball. Two pure states at antipodal points across the sphere are orthogonal pure states (e.g., $\langle 0|1\rangle = 0$, $\langle u|d\rangle = 0$).

One can easily imagine that a point inside the ball could result from an infinite number of possible chords through the ball each with its appropriate probabilities and outer pure state points. This means that the information contained in the density matrix (point ρ) is much less than that of the chord that produced it. The particular knowledge of the pure

[1] Equality results in the equation $\lambda^2 - \lambda = 0 = \lambda(\lambda - 1)$ with eigenvalue solutions $\lambda_1 = +1$ and $\lambda_2 = 0$.

[2] The term "Bloch" sphere (Felix Bloch, 1946) is now often used for qubits and pictured with state $|0\rangle$ or up at north pole, state $|1\rangle$ at south pole, state $(|0\rangle + |1\rangle)/\sqrt{2}$ for x intersection, $(|0\rangle + i|1\rangle)/\sqrt{2}$ for y and no specified radius or location. The Bloch sphere has an earlier relative called the Poincaré sphere dating back to 1892.

[3] This resembles projection operators $P_m = \Sigma|u_i\rangle\langle u_i|$. If we were dealing with Euclidean vectors, we would call this outer product a dyadic (Gibbs, 1884). Its terms would contain unusual things like products of unit vectors, $\hat{i}\hat{j}, \hat{k}\hat{k}$....

[4] $\rho^2 = \rho$ because the middle expression is $\langle\psi|\psi\rangle = 1$ from normalization of psi.

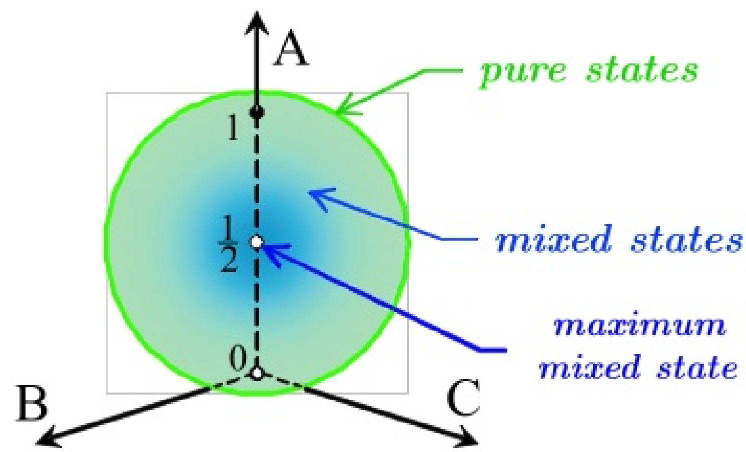


Fig. 1. Geometrical interpretation of the 2×2 density matrix.

FIGURE 1. Bloch 'sphere' B^3 (3-Ball) of density matrices for a 2-state system centered at $\frac{1}{2}I$, from reference [2]. The boundary sphere $\partial B^3 = S^2$ represents "pure" states, while the interior consists of mixtures. The point $A = B = C = 0$ is the south pole of the ball. The properties of B^3 are also discussed in Penrose [3].

states is lost. Still, that density matrix is adequate to calculate the results of experiments, e.g., $\langle A \rangle = Tr(\rho A)$.

2. Examples

To express the density operator in matrix form, we first select a basis $\{|u_m\rangle\}$. Then,

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad \to \quad \rho_{mn} = \langle u_m|\rho|u_n\rangle = \sum_i p_i \langle u_m|\psi_i\rangle\langle\psi_i|u_n\rangle. \tag{2}$$

Rows and columns are labeled by the basis indices. For the unpolarized light example above effectively containing plane polarizations randomly in the H and V directions, we have a 50%- 50% blend of the states $V = \frac{1}{\sqrt{2}}(|R\rangle + |L\rangle)$ and $H = \frac{1}{\sqrt{2}}(|R\rangle - |L\rangle)$ so that

$$\hat{\rho} = \frac{1}{2}|H\rangle\langle H| + \frac{1}{2}|V\rangle\langle V| = \frac{1}{2}|R\rangle\langle R| + \frac{1}{2}|L\rangle\langle L|, \; or \; \rho = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \tag{3}$$

like the example in equation (1). The density matrix is the same whether R,L or H,V is used as a basis. Light only has two polarizations so that some of its math is similar to the case of electron spin one-half (like orientation of a Stern-Gerlach magnet showing spin up or down in a z-direction).

As Penrose emphasizes [3], the above density matrix pertains for all possible orientations such as:
$\hat{\rho} = \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{2}|\downarrow\rangle\langle\downarrow|$, $\hat{\rho} = \frac{1}{2}|\leftarrow\rangle\langle\leftarrow| + \frac{1}{2}|\rightarrow\rangle\langle\rightarrow|$, *or* $\hat{\rho} = \frac{1}{2}|\nwarrow\rangle\langle\nwarrow| + \frac{1}{2}|\searrow\rangle\langle\searrow|$ have the same identical matrix form, ρ. As shown in Figure 1, this density matrix is represented by the central point which is called "maximum mixed."

<u>Trace:</u> Examine Requirement 5 from the introduction: look at $Tr(\rho A)$ knowing that: $Tr(A) = \sum_i \langle i|A|i\rangle$ and $\sum_i |i\rangle\langle i| = I$ at first just for the simplest case $\rho = |\psi\rangle\langle\psi|$ [5]. Then, $Tr(\rho A) = Tr(|\psi\rangle\langle\psi|A) = \sum_i \langle i|\psi\rangle\langle\psi|A|i\rangle = \sum_i \langle\psi|A|i\rangle\langle i|\psi\rangle = \langle\psi|AI|\psi\rangle = \langle\psi|A|\psi\rangle$. So, $\langle A\rangle = Tr(\rho A)$. For the more general mixed state case, $\hat{\rho} = \sum_i p_i|\psi_i\rangle\langle\psi_i|$, we simply have a sum of terms in the calculation.

If we apply the density matrix $\rho = \frac{1}{2}I$ from the example equation (3) onto say a spin-z operator $\hat{S_z} = \frac{\hbar}{2}\sigma_z$, we would obtain $\langle Spin_z\rangle = Tr(\rho S_z) = 0$.

The Pauli 'sigma' matrices are most often presented by:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4}$$

Any 2×2 density operator can be expanded using the Pauli matrices along with the identity, I as:

$$\rho = \frac{1}{2}(I + \vec{a}\cdot\vec{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+a_3 & a_1 - ia_2 \\ a_1 + ia_2 & 1 - a_3 \end{pmatrix}, \tag{5}$$

which obviously has the form of equation (1). The vector $\vec{a} = (a_1, a_2, a_3)$ is called the "Bloch vector" about the central point $\frac{1}{2}I$ of Figure 1. The equation could also be expressed using the hypercomplex quaternions $\mathcal{H} = \{1, q_i = \pm i\sigma_i\}$ (Hamilton, 1843). A maximum mixed density matrix like $\rho = \frac{1}{2}I$ has no distance, $\vec{a} = \vec{(0)}$.

The particular case examples above tend to be boring, so lets now create a partially mixed state. Prepare a merged beam of electrons with spin-up or spin-to-the-right in a $50\% - 50\%$ probability combination. That could be done by combining the output of two Stern-Gerlach magnets with a vertical orientation and a horizontal orientation to give $|u\rangle$ and $|r\rangle = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle)$ while blocking out any down and left spins $|d\rangle$ and $|l\rangle$. Each of the separately prepared spins up and right are pure states. The resulting density operator is now:
$\rho = \frac{1}{2}|u\rangle\langle u| + \frac{1}{2}\cdot\frac{1}{2}(|u\rangle + |d\rangle)(\langle u| + \langle d|) = \frac{3}{4}|u\rangle\langle u| + \frac{1}{4}(|d\rangle\langle u| + |u\rangle\langle d| + |d\rangle\langle d|)$. Then the density matrix is:

$$\rho_{partially\ mixed} = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}, \qquad while\ |r\rangle\langle r| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \tag{6}$$

by itself is a pure state (Block Sphere at the x-axis). $\rho^2 \neq \rho$, so ρ does not represent a pure state. But $(|r\rangle\langle r|)^2 = |r\rangle\langle r|$ which is a pure state. The Bloch vector for ρ is $\vec{a} = (1/4, 0, 1/2)$ or $|a| \simeq 0.56 < 1.0$.

The diagonal elements in a given basis are always the probabilities to be in corresponding states. The off-diagonals measure 'coherence' between any two of the basis states.

3. Entangled States

A large number of copies of the same prepared system is an ensemble state, and density matrices are largely used to describe ensembles (with probabilities measured by frequency distributions). Density matrices can be applied to entangled particles when we have an ensemble of pairs or groups. The most common current way to prepare entangled pairs of photons is using laser beams on a nonlinear crystal. Sometimes an initial photon of some wavelength will split into two photons each having nearly double wavelengths to conserve energy. in SPDC (spontaneous parametric down conversion process) two conical beams are formed where one has vertically polarized photons and the other has horizontally polarized photons. With care about geometry, two divergent rays can show entanglement where a joint state is: $|\Psi\rangle = \frac{1}{\sqrt{2}}(|H\rangle|V\rangle - |V\rangle|H\rangle)$. We let one ray go to system A (often called "Alice") and the other ray go to system B (often called "Bob").

We could consider a state in system A to be labeled $|\psi\rangle_A$ and a state in system B to be $|\phi\rangle_B$. If these two states are independent, then the combined state may be written as a tensor product of the two states in order: $|\Psi\rangle_{AB} = |\psi\rangle_A \otimes |\phi\rangle_B$ [6] (perhaps conveniently written as just $|\psi\rangle|\phi\rangle$). This expression refers to "separable states" or "product states." Joint states are called entangled if they are inseparable (cannot be expressed as simple product states). The HV states from SPDC are an example of entangled states. There is a quantum state given for the system as a whole but its component states cannot be described independently. "There is no way to associate a pure state to the component system A. Alice doesn't know if she will receive an H or a V photon, but once she does know, the state of Bob's photon is immediately determined (as a V or an H). Comparing the results of the two systems will always show perfect correlation (in the absence of noise).

In 1930, Paul Dirac introduced the idea of a "reduced density matrix" as a "partial trace" of the composite density matrix for A over the basis of system B. "The reduced density matrix for an entangled pure ensemble is a mixed state," e.g., $\hat{\rho_A} = \frac{1}{2}(|H\rangle_A\langle H|_A + |V\rangle_A\langle V|_A)$. A necessary and sufficient condition for a bipartite pure state is if it reduced states are mixed. For light, two entangled photons together are a pure state, but each system separately effectively sees unpolarized light.

For the reduced density matrix of A, Susskind [5] says that we 'filter out' Bob's half (or a composite 4×4 matrix) to just get Alice's effective 2×2 matrix. Avoiding operator outer products, the numerical matrix values for Alice are given in his notation by $\rho_{a'a} = \sum_b \psi^*(a,b)\psi(a'b)$, where a and a' are spin states like u,d, and we force Bob's spins to be the same, $b = b'$. For dimension 2 bases of u and d, we have:

$$\rho_{a\,a'} = \psi^*(a,u)\psi(a',u) + \psi^*(a',d)\psi(a',d), \tag{7}$$

$$e.g.,\ \ component\ \ \rho_{du} = \psi^*(d,u)\psi(u,u) + \psi^*(d,d)\psi(u,d).$$

Then for a particular entangled state vector like $|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|ud\rangle + |du\rangle)$, we would obtain the usual maximum mixed reduced density matrix, $\rho_A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$. Note again that this corresponds to the center point of the Bloch sphere.
A and B are highly correlated, but A and B by themselves are random.

4. von Neumann Entropy:

Von Neumann (1932) defined a quantum entropy by a formula similar to a previous classical Gibbs entropy as:

$S(\rho) = -\mathrm{Tr}(\rho\,\mathrm{ln}\rho)$, where ρ is the density matrix. One may use either natural logs or log base 2 ($log_e = \ln$ *or* log_2). If ρ represents a pure state, then its entropy vanishes (the rule $\rho = \rho^2$ implies that $S(\rho) = 0$). Or, we could say that a pure state can always be written in its eigenbase as:

$$(8) \qquad \rho_1 = |\psi\rangle\langle\psi| = col \times row(1\ \ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad So,\ S(\rho_1) = 1\,log(1) + 0\,log(0) = 0.$$

For a finite system, the entropy tells the degree of mixing of the state or the departure from a pure state or the minimum number of bits (log_2) to store the result of a random variable. A student's understanding of this may be initially blocked by the strangeness of the idea of taking the logarithm of a matrix. For square matrices, one can define this as a series expansion of matrices (like that for taking the exponential of a matrix as for Lie groups). And the knowledge that logs and exponentials are inverses can be applied to advantage. But there are also many special tricks for doing it more simply than this. And sometimes, the results are unexpectedly simple (such as $\ln \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$).

For a maximally mixed state such as the example in equation (1) where ρ_2 is the central point of the Bloch sphere, entropy is maximal [8]:

$$(9) \qquad S(\rho_2) = -Tr\left\{\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \log_2 \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\right\} = \frac{1}{2}Tr\left\{\begin{pmatrix} log_2 2 & 0 \\ 0 & log_2 2 \end{pmatrix}\right\} = 1\,.$$

One of the easier tricks for calculation of entropy is the following: Since ρ is a positive semi-definite operator, it has a spectral decomposition such that $\rho = \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|$ where $|\varphi_i\rangle$ are orthonormal vectors, $\lambda_i > 0$ and $\sum \lambda_i = 1$ [7]. Then the entropy of a quantum system with density matrix ρ is the sum: $S = -\sum_i \lambda_i \ln\, \lambda_i = -\,\mathrm{Tr}(\rho \ln \rho)$.

The above can be applied to the case of a bipartite entanglement entropy as the von Neumann entropy of either of its reduced states. That is, for a pure state $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$, it is given by: $\mathcal{S}(\rho_A) = -\mathrm{Tr}[\rho_A \mathrm{log}\rho_A] = -\mathrm{Tr}[\rho_B \mathrm{log}\rho_B] = \mathcal{S}(\rho_B)$ where $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{Tr}_A(\rho_{AB})$ are the reduced density matrices for each partition.

$|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$

The simplest one is the 3-qubit GHZ state: $|\mathrm{GHZ}\rangle = \frac{|000\rangle+|111\rangle}{\sqrt{2}}$. Another important property of the GHZ state is that when we trace over one of the three systems we get $\mathrm{Tr}_3\big((|000\rangle + |111\rangle)(\langle 000| + \langle 111|)\big) = \frac{(|00\rangle\langle 00|+|11\rangle\langle 11|)}{2}$ which is an unentangled mixed state. It has certain two-particle (qubit) correlations, but these are of a classical nature.

CONCURRENCE: can be defined for representing entanglement for two qubit states. $\mathcal{C}(\rho) \equiv \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4)$ in which $\lambda_1, ..., \lambda_4$ are the eigenvalues, in decreasing order, of the Hermitian matrix $R = \sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}$ [6] with $\tilde{\rho} = (\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$ the spin-flipped state of ρ, σ_y a Pauli spin matrix, and the eigenvalues listed in decreasing order. Other formulations[edit] Alternatively, the $\lambda_i's$ represent the square roots of the eigenvalues of the non-Hermitian matrix $\rho\tilde{\rho}$. Note that each λ_i is a non-negative real number. From the concurrence, the entanglement of formation can be calculated.

Monogamy is one of the most fundamental properties of entanglement and can, in its extremal form, be expressed as follows: If two qubits A and B are maximally quantumly correlated they cannot be correlated at all with a third qubit C. In general, there is a trade-off between the amount of entanglement between qubits A and B and the same qubit A and qubit C. This is mathematically expressed by the Coffman-Kundu-Wootters (CKW) monogamy inequality:

$C^2_{AB} + C^2_{AC} \leq C^2_{A(BC)}$, where C_{AB}, C_{AC} are the concurrences between A and B respectively between A and C, while $C_{A(BC)}$ is the concurrence between subsystems A and BC. [cite Quantiki]

It was proved that the above inequality can be extended to the case of n qubits.

If a pure two-qubit state is written as $\Psi\rangle = a|\uparrow\uparrow\rangle + b|\uparrow\downarrow\rangle + c|\downarrow\uparrow\rangle + d|\downarrow\downarrow\rangle$, then concurrence is $C = 2|ad - bc| \geq 0$.

References

[1] StevenWeinberg, "Quantum Mechanics Without State Vectors," arXiv:1405.3483 [quant-ph] 14 May 2014.

[2] Sergey N. Filippov and Vladimir I. Man'ko, "Geometric Interpretation of the density matrix: mixed and entangled states," Journal of Russian Laser Research, Vol. 29, No. 6, 2008,

[3] Roger Penrose, The Road to Reality, Alfred A. Knopf, N.Y., 2004, Chapter 29.

[4] David McMahon, Quantum Mechanics DeMystified, McGraw-Hill, 2006.

[5] Leonard Susskind, Quantum Mechanics, the theoretical minimum, Basic Books, 2014.(e.g., p 196, 211).

[6] Wikipedia on entanglement, http://en.wikipedia.org/wiki/Quantum_entanglement; concurrence.

[7] Wikipedia: e.g., https://en.wikipedia.org/wiki/Density_matrix#Entropy

[8] Dieter Suter, Joachim Stolze, Universitaet Dortmund, Quantum Computing WS13 / SS14, "Measures of Entanglement," https://e3.physik.uni-dortmund.de/~suter/Vorlesung/QIV_WS11/ Entanglement.pdf

5. Appendix

The Schrödinger equation an be written for density matrices, $\hat{\rho}(t)$, as $i\hbar\frac{\partial}{\partial t}\hat{\rho}(t) = [\hat{H}, \hat{\rho}(t)]$.

QUANTUM MEASUREMENT

DAVE PETERSON

ABSTRACT. The concept of quantum measurement poses many problems for students of quantum mechanics. One of the biggest problems is the von Neumann postulate about non-unitary reduction of a wave-packet. But is this really a fact or just a convenient ad-hoc assumption? Quantum measurement is poorly defined in almost all textbooks. [Preliminary].

1. Introduction

Discussions of measurement postulates found in quantum mechanics texts have key statements such as, The state of the system immediately after a measurement is always an eigenvector of an observable with a resulting eigenvalue (e.g., [1]). This is sometimes called the "projection" or "collapse" or "reduction" postulate and is often followed by the comment, "If we perform a second measurement of an observable $\hat{A}$ immediately after the first one (that is, before the system has had time to evolve), we shall always find the same result." The term "eigen" is German for "characteristic."

This postulate can be confusing to students for a number of reasons:

- physics texts are rather notorious for not defining their terms in plain English: How is the word "measurement" defined? What is a "state," what is a system? Instead, we have mathematical statements such as "the state of a quantum mechanical system is described by an element of an abstract vector space" or Hilbert space.

- If one is picturing a state as say a wave-packet, a measurement may collapse the wave so that it no longer exists. So how can we talk about it <u>after</u> it is measured? (e.g., a photon forming a developed spot on a photographic plate). Measuring the position of a photon destroys the photon and hence is not a projective measurement.
- We are drilled to remember that 'A phenomenon is not a phenomenon until it is a measured phenomenon'. We are not supposed to talk about deduced values that have not yet been measured (counterfactuals).
- The equations of quantum mechanics are linear, but "collapse" is non-linear (and random and non-deterministic and irreversible). And the mechanism for this is largely unknown.

Date: May 12, 2015.
email: davepeterson137@gmail.com. Paper updated to 7/7/15.

- A wavefunction may be a linear superposition of different states which should evolve deterministically in time. But measurements always reveal only one particular definite state. How this can happen is largely unknown and is called "The Measurement Problem."
- There are a variety of interpretations of the mathematical formulation of quantum mechanics, and some of them don't believe in collapse or quantum jumping into eigenstates.
- Measurements depend on macro-apparatus which is deemed to be largely classical. But some believe in "quantum all the way up." How do we obtain classical answers?
- Physics texts gloss over the topic of measurement. They don't give many useful examples. They don't say much about state preparation, entanglement with apparatus, possible derivations of the "Born Rule" for output probabilities, history, people, and philosophy.

Can we resolve these problems? First, we can show examples in which a state is projected and preserved for awhile prior to detection and destruction.

2. Some Illustrative Examples of Measurements

The concepts of measurement processes come largely from Johnny von Neumann's 1932 book on quantum mechanics [2], and this became "the Bible of the so-called Copenhagen interpretation of quantum mechanics" [3]. He largely accepted the philosophies of Bohr and Heisenberg who thought of wave functions as representing "our knowledge" of quantum systems (rather than any underlying physics or reality in the wave function itself). That is important because an experiment reveals information, and it is preserved after the measurement (so we can talk about an "after" of "our knowledge"). Also, not many are aware that he was motivated by thoughts about the Compton effect, so we need to talk first about that.

Compton Scattering: The original Compton experiment of 1923 scattered 17 keV x-rays from electrons in the atoms of a carbon target . He observed that the original photon wavelength is increased after scattering as if they were quanta on recoil electrons. This was very important because it convinced skeptics that photon quanta were real. If photons are scattered at an angle, then there is energy-momentum loss from the photon which is transferred to the electron. Detection of the electrons requires that the photons now have specific reduced E and p at some opposing angle. The final photon and electron momenta are required to satisfy conservation of momentum. Then if ${p_e}'$ is measured, we know ${p_\gamma}'$ and could later measure it to be what we deduced. That is, the momentum state $|p_\gamma\rangle$ must have jumped with the interaction (have been projected) prior to later measurement verification [5]. von Neumann then knew that the properties of the yet unobserved partner must have been determined by a first measurement (projected onto the partner). Note that this idea is very similar to what takes place in EPR entangled pairs where a measurement on one party results in a projection on the other. And its values could exist at intermediate times prior to any final measurement.

Single Photon Null Experiments: The Compton example had two particles which interacted with each other. Roger Penrose [4] has simpler one particle examples called "null measurements." His book The Road to Reality uses the symbol **U** for the smooth unitary development of a wavefunction and **R** for state reduction (non-unitary jumping). It is his hope that some future mathematics will include both of these evolutions as limiting cases. His **R** process is also called the von Neumann projection postulate. One example is a photon encountering a 45^o tilted beam splitter (50%/ 50% **BS**) which either transmits a photon or reflects it: $|\psi\rangle = |\tau\rangle + |\rho\rangle$ (rho for reflection). Detectors (apparatus) terminate the end paths of the reflected and transmitted beams. If it is known that a photon goes into the beam splitter but is not detected by the transmission detector, then it is immediately known that its state is the reflected beam (psi is reduced to the reflected state rho, $|\rho\rangle$). Methods now exist for incoming state preparation so that we can know when a photon is actually entering an apparatus. Similarly, if a circularly polarized photon hits a semi-reflected mirror (50%/ 50% **M**), it can be transmitted or reflected backwards. But reflection changes the handedness of the polarization to a new state: $|\psi_+\rangle = |\tau_+\rangle + |\rho_-\rangle$.

Single Photon Polarization: There are various ways to separate out components of polarization of light. One old way is by using a crystal of calcite which has different indices of refraction for different polarizations. A crystal can be oriented so that an ordinary ray passes straight through (say a vertical polarization, V) and deflects a horizontal polarization, H, upwards. Little photocell detectors can be positioned on the resulting two beams (say "h" for a high deflected H beam and "z" for the zero-undeflected V beam. An experiment involves both the beams and the apparatus together, H with h and V with z. A technique called "down-conversion" can convert a high frequency photon into two entangled lower frequency ones, and one of those can be detected and give notice that the other single photon is propagating into an experiment. Then, a triggering of the h detector for the H polarization means that we know there is no photon in the V channel for the z detector. The triggering of one informs the other.

Stern-Gerlach, **SG**: A specially shaped magnet (sharp angle on one side and wide iron face on the other) can give an inhomogeneous magnetic B field (∇B) which can alter the path of little magnetic dipoles passing through the field (e.g., silver atoms with single 5s valence electrons). A neutral silver beam will then split into two beams for spin-up and for spin-down. Atoms with spin $3/2\ \hbar$ will split into four beams with spin $+3/2,\ +1/2,\ -1/2,\ and\ -3/2\,\hbar$. Use of SG deflected beams is one of the most common textbook examples of state preparation. If a silver atom is deflected into an upper beam, then a repeat experiment with another similarly oriented SG magnet will keep it in the upper beam (projections repeat). The spin-up state is preserved over multiple experiments.

A simplest spin-1/2 matrix example is an incoming beam of silver atoms in a superposition $\psi\rangle = a|0\rangle + b|1\rangle$ of spins down and up which ends up in only-down spin states. Let the projection operator be $P_1|\psi\rangle = b|1\rangle$:

$$(1)\quad P_1 = |1\rangle\langle 1| = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\ \textit{and}\ P_1|\psi\rangle = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}$$

A second projection would give $P_1 P_1|\psi\rangle = {P_1}^2|\psi\rangle = P_1|\psi\rangle = b|1\rangle$.

The SG experiment is often used to demonstrate a **projective measurement**. The Feynman Lectures [7] for example discuss "filtered" Stern-Gerlach paths for spin-one atoms in which the lower separated beams are blocked so that only an upper spin-one-up beam passes through. Then a second SG apparatus on that filtered beam will only result in another spin-one-up output beam. The apparatus projects the possible incoming states into just a spin-one-up final state. And since a projection operator has the property that $P\,P = P^2 = P$, a second projection won't change the state.

Unfortunately, in Stern-Gerlach (SG), the strongly inhomogeneous magnetic field is too perturbing to conserve angular momentum (so this commonly used example really doesn't apply to a von Neumann measurement) [5]. But an equation like (1) would still apply to a polarizer on a light beam which is in a superposition of horizontal and vertical input states. And a second similarly oriented polarizer would not change the output light polarization resulting from the first polarizer.

[B] Mathematically, a measurement in quantum mechanics can be considered as a set of measurement operators $\{M_m\}$ over an index of output states. For the simplest case of projective measurements, the measurement operators are the elementary projectors (such as $M_o = |0\rangle\langle 0|$. If $|\psi\rangle = a|0\rangle + b|1\rangle$, then $p_m = \langle\psi|{M_m}^\dagger\, M_m|\psi\rangle = \langle\psi|0\rangle\langle 0|\psi\rangle = |a|^2$. (like in the Born rule). A two-qubit state could have operators like $M_{oo} = |00\rangle\langle 00|$. An observable M can be decomposed as $M = \Sigma\, m p_m$. A projective measurement is repeatable.

The most general kind of measurements are "POVM's" (positive operator valued measure) which can be applied even when we have imperfect measurements that fail. These are called generalized measurements and are non-projective. We create a set of positive operators $E_i > 0$ which sum up to the identity, one ($\Sigma\, E_i = I$).
Example of POVM: Suppose an experimenter is presented with two non-orthogonal states given by: $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. He would be unable to distinguish between them by usual projection methods. What he can do instead is apply POVM elements which at least will never make an error of mis-identification [13][15]:

$$(2)\qquad E_1 = \frac{\sqrt{2}}{1+\sqrt{2}}\,|1\rangle\langle 1|, \quad E_2 = \frac{\sqrt{2}}{1+\sqrt{2}}\,(|0\rangle - |1\rangle)(\,\langle 0| - \langle 1|), \quad E_3 = 1 - E_1 - E_2.$$

An outcome of $E_1 > 0$ implies that the state must have been $|\psi_2\rangle$ because $\langle\psi_1|E_1|\psi_1\rangle = 0$. And $E_2 > 0$ implies state $|\psi_1\rangle$, and $E_3 > 0$ implies "no-inference." Unlike projective measurements, here we have $E_iE_j \neq \delta_{ij}$.

Introductory quantum mechanics tends to deal with closed quantum systems that ideally do not interact with the environment − an external quantum system. This is an unreal abstraction relevant to projective measurements and unitary operators. In thermodynamics, open systems exchange energy and matter with their environment. Similarly, real quantum systems generally do exchange information, energy and entropy with their environments. In these systems, evolution is no longer unitary and the Schrödinger equation becomes inadequate [14].

3. Decoherence

According to Wikepedia definitions, "quantum decoherence is the loss of coherence or ordering of the phase angles between the components of a system in a quantum superposition. One consequence of this dephasing is classical or probabilistically additive behavior." The concept goes back to 1970 in studies by Dieter Zeh. Its main advocate now is Wojciech Zurek who states, "Decoherence selects preferred pointer states that survive interaction with the environment" [10]. His definition is, "Pointer States are the preferred set of states of an open quantum system that are singled out by the persistent monitoring by the environment. They entangle the least with the environment, and are least perturbed by decoherence." He introduced a new name, "einselection" for "environment-induced superselection." In his analysis, he elevates the von Neumann projection postulate to a core postulate, "Immediate repletion of a measurement yields the same outcome.

Examples of measurements include determining a range Delta x for a free particle followed later on by a determination Delta p (Δp). "QND" Quantum Non-demolition measurements avoiding back action. Cavity photons crossed by Rydberg atoms. Von Neumann's non-unitary reduction of the wavefunction eliminates off-diagonal elements from a pure-state density matrix to get a mixed state. But it is possible to avoid this by unitary operators involved in the coupling of a system-detector pair to the environment (e.g., [12]). Decoherence has dissipative and dephasing contributions (modification of populations of states and randomization of relative phases of quantum states).

References

[1] Claude Cohen−Tannoudji, Bernard Diu, Franck Laloë, Quantum Mechanics, Volume One, John Wiley, Hermann, Paris, 1977.

[2] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press, Princeton, N.J., 1955, based on Mathematische Grundlagen der Quantenmechanik, 1932, Springer, Berlin.

[3] Andrei Khrennikov, "Analysis of the role of von Neumann's projection postulate in the canonical scheme of quantum teleportation and main quantum algorithms," 23 May 2008, http://arxiv.org/pdf/0805.3258v2.pdf.

[4] Roger Penrose, The Road to Reality, Alfred Knopf, N.Y., 2005 (e.g., chap. 22, p 527).

[5] Willem M. De Muynick,"Von Neumann-Lüders projection, and its applicability to EPR experiments," http://www.phys.tue.nl/ktn/Wim/vaxjo2008.pdf .
[6] Rieko Sasaki, "Physics of Quantum Information," (course at Stanford, AP 226, Chapter 4, Quantum Measurement. http://web.stanford.edu/*sim* rsasaki/AP226/text4.pdf .
[7] Richard P. Feynman, Robert B. Leighton, Matthew Sands, The Feynman Lectures on Physics, III Quantum Mechanics, Addison-Wesley,1963.
[8] Collapse Theories, Stanford Encyclopedia of Philosophy, 2002 http://plato.stanford.edu/entries/qm-collapse/
[9] James Ward, Philosophy of Physics, 2002, http://www.maths.lse.ac.uk/Personal/james/York/lect4.pdf
[10] Wojciech H. Zurek, "Quantum Darwinism, Decoherence, and the Randomness of Quantum Jumps," arXiv:1412.5206 [quant-ph] 16 Dec 2014. Also see short sketch: PRA 72, 062101 (2005) .
[11] The Information Philosopher site : "The Measurement Problem," http://www.informationphilosopher.com/problems/measurement/
[12] Miguel Orszag, Quantum Optics (Including Noise reduction, trapped ions, quantum trajectories, and decoherence), 2nd Edition, Springer, 1997, 2008 [has Chap 18 on Measurement and Chap. 20 on Decoherence].
[13] Jani-Petri Martikainen, Quantum Information and computing, Lecture 2: Essential quantum mechanics, 2007, University of Helsinki, Finland, http://theory.physics.helsinki.fi/∼kvanttilaskenta/
[14] Davide Girolami, "Quantum Correlations in Information Theory," June, 2013, PhD Thesis Nottingham, http://quantumcorrelations.weebly.com/uploads/6/6/5/5/6655648/girolami.pdf
[15] "Measurement in Quantum Mechanics, C191 Lectures 8 and 9 in Quantum Information and Technology, 9/24/2104, https://inst.eecs.berkeley.edu/∼cs191/fa14/lectures/lecture89.pdf .

Nobel Prize for Topology in Exotic Materials.

David Peterson, 10/9/16 - 11/24/16 - 12/31/16 Revision 1.

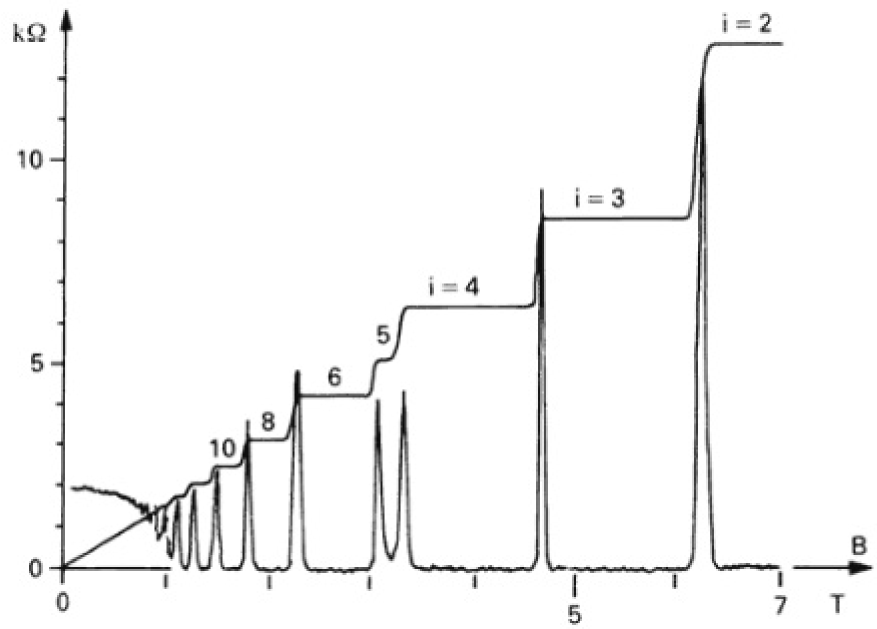
On October 4, 2016, the Nobel Prize in physics went to Thouless, Kosterlitz, and Haldane "for theoretical discoveries of topological phase transitions and topological phases of matter." Half of the prize was given to David Thouless for two key advances: In the early 1970's it was believed that superfluidity and superconductivity were not allowed for very thin 2D layers. Kosterlitz and Thouless showed that wasn't true with the use of topological concepts. And then in the 1980's, Thouless helped explained the mysterious "Integer Quantum Hall Effect" again using topology and "marked the discovery of topological quantum matter." Since then, condensed matter physics of topological materials has blossomed!
[Nobel prizes were awarded on December 10, 2016. But Thouless has not yet presented his work yet. Hopefully he will submit an essay sometime in 2107.]

This note mainly focuses on the **Integer Quantum Hall Effect** ("IQHE" or just QHE) [e.g., see **Figure 1** below]. One author declared in general, "The quantum Hall effect (QHE) is one of the most remarkable condensed-matter phenomena discovered in the second half of the 20th century. It rivals superconductivity in its fundamental significance as a manifestation of quantum mechanics on macroscopic scales." [5] It "is now used to maintain the standard of electrical resistance by metrology laboratories around the world" and measures the fine structure constant alpha accurately to 10^{-8}.

Typically, IQHE needs a two-dimensional electron gas (2DEG), and that can be formed in a thin layer of semiconductor next to an insulator (called an "inversion layer" , e.g., AlGaAs on GaAS). The thickness of this gas may only be 30 angstroms but still can form a broad holistic layer over the relatively large semi-rectangular Hall probe area. The quantum Hall effect is macroscopic! Temperatures < 1 kelvin and magnetic fields > 10 tesla are often also needed [but graphene can show effects at room temperature]. Applied voltages in the long x direction of a rectangle cause a build-up of voltage in the y width direction, so conductivity technically needs be a 2D tensor: $J_i = \sigma_{ij} E_j$ $(includes\ J_x = \sigma_{xy} E_y)$. Hall resistance is measured in the cross y direction and was observed to change in integer steps on plateaus .
$\sigma_{xy} = \nu e^2/h$ or $\varrho_{xy} = -h/\nu e^2 = R_K/\nu$. $R_K = h/e^2 \sim 25.6\ k\Omega$ is called the "von Klitzing constant (and is good to 9 figures). [note that the fine structure constant is
$\alpha = e^2/4\pi\varepsilon_o \hbar c$ $[SI]$, $so\ R_K\ determines\ \alpha$]. A requirement for topological integer plateaus is having imperfect materials (doping ions, surface roughness, random disorder -- and most materials do have uncontrolled imperfections). Large magnetic fields are needed to see the biggest "ground" plateau. And, note that (with the right setup) going to 30 T may introduce an unexpected "fractional plateau" (1/3rd) -- a separate and very weird arena with largely different physics (see Fractional Quantum Hall Effect FQE in a section below).

Details of the IQHE are intricate, dovetail in an almost conspiratorial way, and are difficult to the point of first requiring reading an entire book on the subject (such as that of David Tong, [1]). Robert Laughlin (Nobel 1998) would insist that this new physics is "emergent" from collective phenomena and can not be mathematically deduced from fundamental physics (the whole is greater than its parts). Others will try anyway but with some mystery and opaqueness.

Before a more detailed view of all this can be discussed, it is necessary to first introduce several preliminary topics: the standard Hall effect in classical physics, Topology, Landau levels, Anderson Localization, Fermi levels, and Edge modes.



[Source: K. v. Klitzing, G. Dorda, M. Pepper: *New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance*, Phys. Rev. Lett. **45** (1980) 494-497, Figure 1.]

Figure 1. IQHE Plateaus shown by quantized electrical resistance Rxy versus applied magnetic field B and labeled by Landau level integers, i. Continuing B to above 10 T would reveal the i=1 plateau. The Landau Level (LL) spikes are for direct lengthwise resistance Rxx > 0 and occur at the transitions between plateaus. On the plateaus, Rxx = 0.

Classical Hall Effect:

Every freshman physics text presents the classical Hall effect of 1879. If a current, I_x , flows through a thin metallic strip that has a strong perpendicular magnetic field, B, going through it, then a potential difference develops between the sides. Some density of charge carriers, n, in the strip flows with a slow drift speed v and experiences a cross field force

$F_y = qv_x \times B_z$ *with current* $I_x = ne\delta w v_x$ where delta is thickness (very thin) and w is cross width in the y-direction. Then,

$$F_y = ev_xB = BI_x/new\delta. \text{ But } V = Fw, \text{ so, } V_y = I_xB_z/ne\delta.$$

$$\text{And resistance } R_{xy} = V_y/I_x = B/ne\delta \qquad \textbf{Eqn. 1}$$

The formula says that even to get micro-volts of voltage will require high magnetic fields (like $B \geq 1\ T$ where tesla = 10,000 gauss in current college labs) and extreme thinness ($\delta \sim 20$ microns or less).

The American physicist Edwin Hall used thin gold leaf for his conducting strip and revealed the effect well <u>before</u> the discovery of the electron. So what he revealed was that the quantity “ne” flowed through the gold as a negative current. Positive current flow would give an opposite side-to-side voltage (good for semiconductor positive hole flow). Using E as the induced electric field sideways and J is the current flow density through the strip, a “Hall <u>coefficient</u>” was defined as: $"R_H" = E_y/J_xB_z = V_y\delta/I_xB = -1/ne$ **[Eqn. 2]** showing a way to measure carrier density or magnetic field B (“Hall effect probe). Note that this unfortunate naming convention is different from the Hall <u>resistance</u> above, so resistance is $R_{xy} = B\,R_H/\delta$. The rewards of this classical measurement are knowledge of charge density for carriers and resistivities for materials. And in the 20th century, carriers could be “holes” with an effective positive charge. One should also study the “Drude Model” which adds a friction term to cyclotron motion in the form of a scattering time, τ. It is this model that makes clear that conductivity should be treated as a 2x2 tensor leading to resistivities: $\varrho_{xx} = m_e/ne^2\tau$ *versus the usual* $\varrho_{xy} = B/ne$. And when we find that $\varrho_{xy} \neq 0$, *then* $\varrho_{xx} = 0 \Longrightarrow \sigma_{xx} = 0$! (unexpectedly the system is then a perfect insulator). The longitudinal Rxx depends on sample composition and sample length.

<u>Topology:</u>

We tend to think of topology as the counting of “holes” through geometric objects (something through which a string can thread); and for one hole, we consider a coffee cup to be “the same as” a donut. The number of holes represent “topological invariants” that are usually integers. But, the term “hole” can also apply to objects of any dimension. So, for example, the inside of a sphere is called a 2-hole (something that can be filled with water). There are also a variety of types of topological indexes and other concepts that are hard to picture.

One goal of topology is to identify properties of objects that are invariant under continuous deformations. A simplest example of a topological concept is that of “deformation classes” or “path components” of geometric regions, S. This means that for any two points, there can be a continuous path ending on the points, and this idea obviously applies to a 2-sphere, or a torus surface, or infinite Euclidean spaces E^n. The symbol $\pi_o(S) = 0$ is used for the set of all path segments that can be deformed into each other. A virtue is that "global topological properties are robust against local perturbations [7]."

But, if there is a forbidden "**gap**" separating two materials, then there is no continuous path between them. The idea of a forbidden barrier also applies to physical "phases" so that solid ice is separate from water fluid (liquid/gas) on a pressure versus temperature plot (a path does exist between liquid and steam by going around the "triple point"). There is a "phase transition" between between solid and fluid states. We now know that there are other types of phases in condensed matter physics such as topological superconductors, topological insulators, superfluidity, and now the quantum Hall states. In IQHE, there is a phase transition at specific energy levels so that a normally insulating material suddenly becomes a good conductor.

The role of topology in condensed matter physics often enters through quasi-momentum on the **"Brillouin torus."** For crystals, electron states depend on the geometry of the lattice which generally repeats from atom to atom. The potential energy is periodic like the lattice, and the wavefunction is also periodic: $u(x + na,\ y + nb) = u(x, y)$ for a rectangular lattice. The primary difficulty is dealing with the vast variety of possible types of crystal structures. Including momentum gives a "Bloch wave:" $\psi(r) = e^{ik \cdot r}\, u(r)$ where k is the crystal wave vector and momentum $p = \hbar k$.

The simplest rectangular physical lattice has another view called the "reciprocal" lattice with primitive cell sides: $A = 2\pi/a \text{ and } B = 2\pi/b$ which is effectively a Fourier transform of a simple physical lattice. Reciprocal lattice points or vectors G in this Fourier space are $G = hA + jB$ where h and j are integers. Crystal wave diffractions are satisfied when $\Delta k = G$. A cell of size A x B is called a "first Brillouin zone." Because of periodicity, the opposite sides are "identified", and that means homeomorphic to a <u>torus</u> ($T^2 \text{ for } 2d \text{ and } T^3 \text{ for } 3d.$). "The fundamental group" for the torus is: $\pi_1(T^2) = \pi_1(S^1) \times \pi_1(S^1) = Z \times Z$, where Z is the set of integers (e.g., representing "winding numbers" about a circle).

This means that there could be non-contractible <u>loops</u> (rather than the previously mentioned arcs) around the torus representing many integers of winding numbers. [Note that S being "simply connected" implies that $\pi_o(S) = 0 \text{ and } \pi_1(S) = 0$]. "The full ensemble of states over the Brillouin torus is always trivial." But an energy gap can cause a split into two well separated sub-ensembles each with non-trivial topology. This is related to Thouless' original Chern topological index. "The Chern number is topological in the sense that it is invariant under small deformations of the Hamiltonian." [21]

As a short hint on these topics: Chern number, Berry phase, and classical "Gauss-Bonnet" Euler characteristic can all be calculated as integrals.

$\chi(sphere) = (1/2\pi)\int_M K dM = \chi = 2 - 2g.\ E.g.,\ \chi(S^2) = 2,\ g = 1 - \chi/2 = 0,\ and\ \chi(T^2) = 0$ [23].

K is the "curvature" of a Manifold, g = genus = holes/handles for a 3D surface. A sphere has no handles and a torus has one hole. A simpler example is for a 2D circle

Circle S^1 : $(1/2\pi)\int_{circle} K ds$ = $(1/r)(1/2\pi)(Cir = 2\pi r)$ = $1 = \chi$ $in\ 2D$.

The 3D genus and Euler characteristic also pertains to the old high-school geometry: vertices - edges + faces= V-E+F. For a 4-faced tetrahedron, 4-6+4 = 2 so g = 0 (no holes).

Berry Phase uses Stokes' theorem to get a form: $\gamma = \int_S dS \cdot \Omega(R)$ where Omega is a Berry curvature from a Berry connection and R is a vector parameter of time.

The topologically invariant Chern number, $Ch_n\ or\ c$, comes from the integration of "Berry curvature." A nonzero Chern number says that there is an obstruction in applying Stokes theorem over the entire parameter space [-- see "Geometry in Modern Physics" [6]]. If one wants to see plentiful applications of topology, condensed matter physics is the place to be -- however, the dovetailing of the Chern numbers to IQHE is acknowledged to quite difficult [9].

Many articles on topology and physics deal with the "real world." But, the topology in the quantum Hall effect is really a topology in a quantum state" and quantum topology is now used for many application. "Berry phase is the simplest demonstration of how geometry and topology can energy from quantum mechanics" and at the heart of the IQHE. This phase shift occurs when a complete loop is made in some parameter space and is geometric and separate from the usual Edt and kdx phase contributions. The leading example is the: Aharonov-Bohm (AB) effect with phase change $\gamma = \oint_C eA_i dx^i$ (e.g., for a closed path around a solenoid). And this is applied below.

In modern condensed matter experiments, one can additionally see analogue cases of formation of Dirac monopoles and also Yang monopoles with non-vanishing 2nd Chern number measured for the first time [7]. A research article by NIST said: "Fundamentally, topological order is generated by singularities called topological defects in extended spaces, and is quantified in terms of Chern numbers, each of which measures different sorts of fields traversing surfaces enclosing these topological singularities. Here, inspired by high energy theories, we describe our synthesis and characterization of a singularity present in non-Abelian gauge theories - a Yang monopole - using atomic Bose-Einstein condensates …"

Topological materials have topological properties that are "robust and insensitive to perturbations and impurities." They "stay the same if you continuously change the system: stretching it, straining it, shaving off some layers – or really any change that doesn't cause a phase transition." [17]

Claimed definitive explanations of IQHE can be shown in several different ways; and one seems to require "Non-Commutative Geometry," [Bellissard, 1994, ref. [3]]. Hall conductance is a non-commutative Chern number, "Ch." That is, $\sigma_{xy} = \nu e^2/\hbar = (e^2/h)\, Ch(P_F)$ interpreted as a Chern character from a "Kubo-Chern" relation. The first inroad to understanding IQHE quantization was given in a famous (*-ly undreadable*) 1982 paper referred to as "TKNN" [8] for its four authors (one of them being David Thouless). It says “Hall conductance is quantized whenever the Fermi energy lies in an energy gap, even if the gap lies within a Landau level.”

Landau Levels:

A first step is to talk about electron motion in a thin film with a normal magnetic field, B. The Lorentz force F = qvB will be balanced out by “centrifugal” force $F = mv^2/r$ where $v/r = \omega = \mathit{angular\ motion}$. The electron will go around in circles with a “cyclotron” frequency” $\omega_c = qB/m$ where m is the effective mass of the electron. Since Boltzmann’s constant is $k_B = 8.61 \times 10^{-5}\ eV/K$ and lab temperatures are below 1 K, thermal fluctuations have negligible effect. This allows for the emergence of quantum effects such as quantized Landau levels and quantized magnetic flux. A typical energy for $\hbar\omega_c \sim 10\ meV\ for\ fields\ B \sim 10\ T$. An analogy with old Bohr, one aspect of circular motion is that a circumference has to be integer multiples of wavelengths round the circle.

Mathematical Derivations:

The presence of a magnetic field in a z-direction alters a term in the Hamiltonian as $H = (p - qA)^2/2m$ as if a vector potential A times charge acted as “electromagnetic momentum” The term (p-eA) is called “canonical momentum,” as opposed to usual “mechanical momentum” $p^{\mu}{}_{mech} = m\,\dot{x}^2$. The vector potential is not gauge invariant, and Lev **Landau** picked a special “Landau gauge” for A: $A_x = -By$ (or alternately $A_y = Bx$ with all other $A_i = 0$) which acts as a simple shearing field indeed giving $\nabla \times A = B$ as it should. Then the Hamiltonian could be written as $H = [(p_x + eBy)^2 + p_y{}^2 + p_z{}^2]/2m$. If we were to label an “offset” distance as $y_o = -p_x/eB$, we could write out a term, $[m\omega_c{}^2(y - y_o)^2/2]$, exactly matching the first term above (one has to expand both squares and match up the terms). The this second degree of freedom is the coordinate of the center of the cyclotron orbit.

Now, the standard “Linear Harmonic Oscillator” **(LHO)** has a similar form $H = p^2/2m + m\omega_c{}^2y^2/2$ where the last term incorporates a vibrating spring energy. For IQHE, we have a $(y - y_o)^2$ term instead of a y^2 term implying a new off-centering concept. This displacement can be thought of as $y_o = kl^2$ where l is “magnetic length” $l = \sqrt{\hbar c/eB} = 25.7nm/\sqrt{B\ teslas}$. In the IQHE, B includes many magnetic flux quanta $\Phi_o = h/e \sim 4 \times 10^{-15}\ Wb$ -- webers a unit of magnetic flux (or half that value for the case of Cooper pairs for superconductivity vortices) so that the density of magnetic flux is $B = \Phi_o/2\pi\, l^2$.

Using these Hamiltonians in a quantum mechanics setting requires solving the Schrodinger equation where H is treated as an operator: $\hat{H}\Psi = E\Psi$. We don’t have to do that

here because all standard texts solve the easier LHO problem and present its wavefunctions. We then know that the quantum linear harmonic oscillator ends up having quantized energy levels according to the famous formula: $E = (\nu + 1/2)\hbar\omega$; and because the Hamiltonians are similar, that will also apply to the energies for circular motion Landau Levels. So energy could be pictured as increasing in steps of
$\nu = 0,\ E_o = \hbar\omega_c/2,\ \ and\ then\ \nu = 1,\ E_1 = 3\hbar\omega_c/2,\ and\ \ \nu = 2...\ So\ \Delta E = \hbar\omega_c \sim 10\ meV$ is the gap separation energy.

So, electrons may ideally only occupy orbits with discrete energy values.
And, the n above determines the integer n in the IQHE! The Landau level location are where the IQHE makes its ϱ_{xy} jumps in cross resistivity, and the "spikes" in Figure 1 represent direct resistivity ϱ_{xx}. These are also peaks where the Landau "density of states" [DOS = g(E)] or "degeneracy" is high. The strangest result is the occurrence of a "phase transition" of extended states at every Landau Lever band center (i.e., the "spikes").

Note that the energy here didn't depend on the $p_x = \hbar k_x$, so degeneracies can exist. If LHO eigenstates are labeled by $|\varphi_n>$, then the state of an electron can be:
$\Psi(x, y) = exp(ik_x x)\ \varphi_n(y - y_o)$ which depends on the quantum numbers n and k_x. As the n values and energy levels rise, it turns out that the now fuzzy wavefunctions increase in radial size as well [as $< r^2 > = 2(n+1)\hbar/eB$ (wider circles). And they also have angular momentum: $L_z\Psi_n = \hbar n\ \Psi_n$. This radial increase turns out to be important to the understanding of IQHE.

As mentioned before, these Landau levels can only be observed for very low temperatures and very strong magnetic fields: $\hbar\omega_c >> kT$. It is important to estimate how many sublevels can exist in a Landau level (the degeneracy of the ground state). The answer is $N \sim BL_xL_y/\Phi o$, where L is the width of the Hall strip [5] and Phi is a tiny quantum of magnetic flux. If due to Zeeman energy splitting, it is"typically about 70 times smaller than the cyclotron energy" [9] for GaAs. The degeneracy increases with the applied magnetic field through a characteristic area. <u>"There is one electron-state per Landau level per flux quantum."</u> So, in tests where the B field ramps up, more electrons can go into the lower LL's. That is why the high B fields of Figure 1 reveal the low labels of the LL's.
Levels are characterized by integer called "filling factors," $\nu_f = hn/eB$ where n is the surface electron density and ν_f is "the ratio between the total number of electrons and the number of states on one Landau level." [17]

<u>**"Anderson Localization:"**</u>

In general, "electronic conductivity should be directly proportional to the electron mean free path [4] which is typically ~ 100 nm. But, in 1958, Philip Anderson wrote a complicated paper suggesting that electron scattering can be much more localized in the presence of many crystal defects. Doped semiconductors is one example of a disordered crystal lattice (acting somewhat like random potentials at crystal sites). In Anderson's electron localization, the electron zigzags between impurities resulting in a smaller mean free path and hence greater resistance. If a "localization length" is labeled as ξ, *then* $|\psi(r)|^2 \sim e^{-|r|/\xi}$. A short localization length restricts electron propagation. If motion is free across the entire Hall strip, then probability is unlocalized or "extended" and constant. In the presence of large B fields, localization is different; and there is only one critical energy allowing for an extended state (pretty much in the center of a DOS peak at Landau energy). Disorder broadens the DOS peaks, and anything to the sides of dead-center still is localized with only the middle being delocalized (a strange emergent result that is hard to understand in any simple way).

Impurity scattering dominates at very low temperatures. It happens that localization lengths diverge exactly at Landau levels whereas in-between these levels, direct conductivity vanishes and Hall body electrons are localized. That means that the plateaux in Figure 1 owe their existence to localization from crystal disorder. Modeling of the effects of impurities can be accomplished by using a random potential V(x) in the electron Hamiltonian [1]. Quantized resistivity persists on these precise plateaux over a range of increasing magnetic field strength, B, and charge carrier density, n.

The details of LL conductivity are very tricky and subtle. Between two adjacent Landau energy levels, there is strong Anderson localization; and localization blocks conductivity. Bulk states are insulating. Exactly at the Landau level, the localization length diverges into conductive "extended states." As one increases electron density at a Landau level, the filling gets added into the bulk localized states caused by disorder so that they don't add on to net transport (Hall conductivity is a quantized constant > 0). The conductivity $\sigma_{xy} = \nu e^2/h$ gets "stuck." In-between Landau levels, increasing the Fermi level only occupies localized bulk states. Only the narrow centers of the Landau Levels (LL's) have current carrying extended states. $\sigma_{xx} \to 0$ *and* $\sigma_{xy} > 0$, *then* $\varrho_{xx} = \sigma_{xx}/(\sigma_{xy}{}^2 + \sigma_{xx}{}^2) = 0$, zero direct resistivity too.

Summarizing the above:

Magnify a little part of Figure **1** to consider just one of the plateaus between a direct "spike" on the left and another spike on the right. The spike itself results from a sudden increase of "localization length" or "extended state" phase change from insulator to metal allowing a boost in conductivity so that $\xi >> 0$, $\sigma_{xx} \sim h/e^2 > 0$ *along with* ϱ_{xx} *and* $R_{xx} > 0$. In the plateau we have the emergence of fixed (stuck, persistent, quantized) non-zero resistivity and conductivity for topological invariants ϱ_{xy} *and* σ_{xy} *but also* σ_{xx}, ϱ_{xx} *and* $R_{xx} \sim 0$. *And* $\xi \sim 0$ means strong Anderson localization.

There are now many approaches to the physics of localization including some that treat it as a critical phenomenon using a size varying "scaling function $\beta(g)$" -- as in quantum field theory (QFT). In 1984, Libby, Levine and Pruisken attacked the phase change problem incorporating a "theta angle" into the Anderson model [8]. This is an idea of an "instanton vacuum" and "nonlinear sigma model" borrowed from quantum chromodynamics (QCD) for quark confinement versus deconfinement. Then there is a "renormalization" flow diverging at the Landau energies and producing quantization. This means The robustness of IQHE plateaus is seen as a large scale emergence. It is rather amazing that ideas from high energy physics may pertain to solid state physics, but they are gathering experimental validation [11]. But also recall that some of these particle physics concepts originally came from Anderson's studies in solid state physics (e.g., the Higgs Symmetry Breaking idea). Unfortunately, Pruisken's field theory is qualitative and has not been able to calculate quantitative results. Numerical approaches then seem best, and the fluctuations seem to be multi-fractal in nature.

The insulator to metal transition looks like a critical point phenomenon of the form: $\xi/\xi_o = |E_o/(E - E_c)|^{2.33}$ *where* $\xi_o \sim$ *magnetic length*, $E_c =$ *critical pt. LL*, $E_o =$ *characteristic* energy . The power drop-off $\nu \simeq 2.33$ is a universal constant. Despite this blow-up to infinite delocalization, longitudinal conductivity is still finite e.g., $\sigma_{xx} \sim 0.54\, e^2/h$. The IQHE phase change is one of the best known examples of a quantum critical point of a disordered system,. In this case, it is a continuous phase transition or second order phase transition with zero latent heat [12].

Fermi Level:

Electrons are half integer spin fermions obeying the Pauli exclusion principle. That means that two electrons with the same quantum numbers cannot get too close to each other. The number of states per unit volume with a given energy ε_i *(electron volts eV)* and degeneracy $g_i(\varepsilon_i)$ is given by $N_i = F(\varepsilon_i)g(\varepsilon_i) = g(\varepsilon_i)/[1 + exp[(\varepsilon_i - \mu)/kT]$, *where* $F(\varepsilon_i)$ is called the Fermi-Dirac distribution, and mu is "chemical potential." The term "Fermi energy" usually refers to "the (kinetic) energy difference between the highest and lowest occupied single-particle states in a quantum system of non-interacting fermions defined as *always at an* absolute zero temperature." In a metal, the term "lowest occupied state" usually means the bottom of the conduction band.

The "Fermi level" or "electrochemical potential" in a metal at absolute zero is the energy of the highest occupied single particle state including both kinetic and potential energy (the energy of the lowest state). It is the surface of the sea of electrons such that no single electron can rise above it. So, the Fermi level is the total chemical potential for work required to add one electron to the body.

In solid state theory, atoms are packed close together so that their previous discrete energy levels merge into a band of energies such as the valence band. In semiconductors, there is an energy gap between a valence band and higher conduction band and the Fermi level lies in the forbidden gap. In metals, there is no gap and electrons can move freely (conduct). An insulator means having a large gap (no free conducting electrons). With temperature added, thermal energy can excite electrons in a band and the Fermi level can be set at an average occupancy of 0.5 . So some semiconductor electrons can jump up to the conduction band leaving holes in the valence band. Near absolute zero, electrons fill to the Fermi level with a number of sub-bands below it depending on the applied B field.

In IQHE, increasing the B field increases the degeneracy of each LL. That means that the Fermi level will fall with increasing B field. When the Fermi level lies between Landau energy levels, then all lower Landau levels will be filled. Or we could say that a decreasing B implies that each LL holds fewer electrons and the Fermi energy will go up. “But rather than jumping up to the next Landau level, we now begin to populate the localized states. Since these states can’t contribute to the current, the conductivity doesn’t change. This leads to exactly the kind of plateaux that are observed with constant conductivities over a range of magnetic field” [9]. There is a strange conspiracy that the “current carried by the extended states increases to compensate for the lack of current transported by the localized states. This ensures that the resistivity remains quantized…” [9].

Edge Potential and currents:

Circular motion of electrons is geometrically blocked at the side edges of a thin Hall strip. Essentially, the electron performs half a circle there, bounces back and executes another sequential half circle. This is called “skipping motion” in which electrons can only move in one direction and cannot backscatter from impurities. The net result is a dissipationless edge current flowing forward on one side and flowing backwards on the opposite side [1] (chiral motion). This persistent circulating current is real and measurable. Potential V(x) is highest at these edges, and the edge material acts as a metal. The Landau levels are pushed up at the edges and can rise above the Fermi level. But the bulk in-between is more like an insulator. Impurity scattering is low at these edges, but yet impurities are important for the emergence of the Hall plateaux [1]. The population of edge states traverses the band gap between the valence band and conduction.

On an energy diagram E versus distance across a Hall strip ($0 \leq y \leq W$), each Landau level has a “bathtub” shape (flat on the bottom and rising strongly in energy at the edges). For a given Fermi level, several of these bulk LLs may lie below that level. For example, at plateau i = 2 may have LL n = 0 and n = 1 lying below it. The LL extended states crossing the Fermi energy level correspond to the transitions between plateaus (the “edge states”). Some sources suggest that direct current may be “carried entirely by the edge states.” With high B fields, the electrons that carry current are confined to the edges by the Lorentz force, one for each LL.

When a y- potential difference is introduced across the width, more electrons are introduced across the width and accumulate more on one side than the other -- the bathtub is tilted towards one side. The fermi potential is the same on both sides. Hall voltage gives the Hall conductivity $\sigma_{xy} = I_x/V_H = e^2/h$ [1] (and the appearance that current is carried by the edge states). So, the bulk of the electron gas is an insulator, but along its edge, electrons circulate as an example of the quantization of Berry’s phases [22]. This is related to the concept of “topological insulators” with conducting edge states where “spins of opposite sign counter-propagate along the edges.” (quantum spin Hall [QSH] states)

The most important observable in IQHE is that cross-conductivity is quantized. But if a cross voltage has been built up at equilibrium, why should there still be any current? The answer is that there is always current at the edges of the Hall width, and current in-between can flow from edge to edge. That flow may be incremental widthwise from one LL state to a neighbor and then on to an edge.

Integer Quantum Hall Effect:

The Quantum Hall state is the simplest example of a topologically ordered state and occurs for an electron gas in two dimensions. The Hall conductivity changes stepwise with increasing magnetic field. But, for ultra thin and ultra cold samples, the physics becomes quantum mechanical and crosswise Hall conductance can change by integer steps! $\sigma_{xy} = \nu\, e^2/h$. This is the Integer Quantum Hall Effect (IQHE). The steps or plateaus have incredibly precise values enabling ultra-fine electrical measurements.
von Klitzing [2] in 1980 was the first to discover that conductivity here was exactly quantized and won a Nobel prize in 1985) [again see Figure 1].

A big question is "Why do the steps change by integer multiples?" and "Why are the plateaus broad?" rather than changing with magnetic field like the Hall formula. Thouless helped provide answers to these questions. The plateaus are broad and stable due to Anderson localization between quantized Landau energies. These plateaus exist "when the Fermi energy crosses an extended state level." Why the conductance changes by integer multiples is given by advanced topology arguments utilizing Chern theory such as that in the TKNN formula. The IQHE conductance is robust because it is a topological invariant of the system immune to deformations [9]. A plateau means that the delocalized sub-bands are completely filled. The conduction electrons cannot jump from one energy level to another, since there are no available energy levels for them. As a result, the scattering of conduction electrons, with loss of energy, cannot happen." [17]

Attempts to model Quantum Hall transitions included an early use of semi-classical percolation and quantum tunneling. This is still sometimes used but no longer stressed. Delocalization may now be discussed using Topological Field Theory [wikipedia]. There is something mysterious about half-filled Landau levels that makes them special and suddenly metallic. No theory fully explains why the quantization is so perfect and unaffected by the geometry and purity of the material [21].

"Laughlin Gauge Argument":

Most explanations of Hall quantization are advanced and difficult. The first explanation is the simplest and most referenced [13] -- but still tricky. In 1981, Laughlin considered a 2D rectangular metal strip of length L and width W bent into a circle and also having a normal magnetic field Ho everywhere on the loop (e.g., from an imaginary magnetic monopole). He considers the "disordered case with the Fermi level in a mobility gap.." Let there be a current I resulting in voltage drop V across the width by the Lorentz force. He then considers what happens when magnetic flux is introduced down through the middle of the circle (where magnetic flux is defined as the field times the cross sectional area). For this we need to first look at the Aharonov-Bohm (AB) effect of the vector potential A on electron phase. A is important because of canonical momentum in the Hamiltonian: $H = (p - eA)^2/2m + eE_o y$. There is a field B inside the solenoid of radius R but no magnetic field outside, just a vector potential field. For this A field around the outside of a solenoid (or uniform A field around the ring in this case) at radius rho:

$A_{out} = B_o R^2/2\varrho = flux/circumference = \varphi/L$, Then the AB phase change will be:

$eA\Delta x/\hbar = eAL/\hbar = e\varphi/\hbar.$

If we insist that the phase around the ring be a single valued function (rather than a multivalued winding function) then the total circle phase change must be integer multiples of 2 pi. So,

$AB\ phase = n2\pi = 2\pi eAL/h,\ or\ A = nh/eL\ for\ extended\ states\ (or\ n = ALe/h = \varphi\, e/h).$

Now add one magnetic flux quantum, $\Phi = h/e$ h/e (so delta n=1). Laughlin says that this sort of gauge invariance requirement maps the system back into itself.

This is an interesting result, that one magnetic flux quantum changes the AB phase by one wavelength around L.

Now notice that power = dU/dt = VI, but $V_x = \oint_L E_x dx = -d\varphi/dt,\ so\ I = dU/d\varphi = dU/LdA$.

Laughlin then claims that one electron per LL is transferred from one edge of the strip to the other edge by ratcheting in successive stages across the width (shift register). This shifting is related to the magnetic lengths and y's discussed above under "Landau Levels" . Current flow in the x direction drives a voltage in the y direction. This current is the transfer of n electrons across the width so that

$I_y = ne/\Delta t = neV_y/\Delta\varphi = ne^2/h$! (using the dt from Faraday's law above).

So, Hall current in the y direction is quantized.

[Of course, there are some assumptions and details left out and still to be addressed, as they are in references [12] [13]]. He adds, "At the edges of the ribbon, the effective gap collapses and communication between the extended states and the local Fermi level is reestablished."

Many articles present the above argument as a "Corbino Annulus" instead of a ring. This model originated in a 1911 study on magnetoresistance. Insertion of central flux then causes migration of charge from the inside radius to the outside.

Fractional Quantum Hall Effect (FQH).

Beyond the Integer QHE: In 1982, Stormer and Tsui first discovered a new quantum Hall effect showing that the ratio of electrons to magnetic flux quanta can occur in p/q integers like ⅓ or ⅖ ! Particles can act as if they had a fraction of the charge on the electron. This is a new state of matter. Remember from above that the IQHE identified one electron state to a Landau level and a magnetic flux quanta. In general, the microscopic origin of the FQH remains unknown, a big work in progress. But Laughlin presented reasoning for the special case of a 1/q state and eventually won a Nobel prize (along with Tsui and Störmer). The FQH requires a "many-electron wave function" (like the 1983 Laughlin example) resulting in fractionally charged "quasiparticles." This is a type of Bose-Einstein condensate in which electrons are bound with an odd number of vortices which can have neighboring depleted charge regions leading to effectively fractional charge.

Resulting composites may be "anyons" that are neither fermions nor bosons. This is dominant in FQE theory, but no anyon has been conclusively seen experimentally. If they do indeed exist, the FQH is the place to find them. The IQHE depends on absence of electron-to-electron interaction, but the FQH depends on it and wants smoother surfaces. The vast number of fractional FQE bands currently requires doing experiment first and trying to formulate theory patterns second. IQHE and FQH are examples of emergent collective order supposedly not deducible from fundamental physics but only from experiment. This follows the new philosophy of Philip Anderson's "More is Different" and Robert Laughlin's "The end of reductionism." The FQH phenomenon are very similar to IQHE except for the transfer of fractional quantum numbers.

FQE is an example of "topological order" with patterns of long-range entanglements, and the changing from pattern to pattern requires a phase transition. This concept lies beyond that of topological insulators, topological superconductors, and traditional Landau symmetry breaking. It may also include high temperature superconductivity and also the IQHE above with a "Chern number of the filled energy band." FQE has Chern-Simons gauge theories as their effective low energy theory. Topological order has "quantized non-Abelian geometric phases of degenerate ground states." (Wikepedia).

For the IQHE, we depend on material disorder. But FQE needs minimal disorder (cleaner samples) to show its fractional value plateaus.

Kosterlitz-Thouless (KT) Transition: Earlier Work.

Before 1960, it was believed that long range order in two dimensional solids was impossible. In the 1970's, a new "topological order" was discovered in which 2D vortices and anti-vortices (which are not whirlpools) pair together allowing unexpected 2D superfluidity and superconductivity. A 1972 "KT" paper was titled, "Long range order and metastability in two dimensional solids and superfluids" [15]. The authors first considered standard dislocation theory and the pairing of "up and down" dislocations but noticed that their observations should also pertain to vortices in superfluids as well. At low temperatures, pairs of "opposite" dislocations pair up closely, but at high temperatures they freely separate and allow a viscous response. They studied what is called the XY model (2D classical rotor or spin model) on a 2D lattice. The KT transition lies between high temperature direction correlations (which decay exponentially fast) and power-law low temperature decay. A Russian, Vadim Berezinskii, did similar work resulting in the name “BKT transition.” It was noted that superfluid vortices can form above a critical temperature but not below it. Or, vortices and anti-vortices are free above a critical temperature but paired very close below it. This is a collective phase field unbinding effect that is universal in variables regardless of the chosen system being studied and correlation lengths diverge exponentially [15]. Again, renormalization group equations seem to apply.

A KT transition has been confirmed experimentally in proximity-coupled Josephson junction arrays, and “quasi-long range order” has been applied to thin films of superfluid helium, thin-film superconductors, and other systems.

Duncan Haldane:

Duncan Haldane is a British physicist who did his initial work on one-dimensional chains, and 1D seems less glamorous than the 2D electron gas problems discussed above. In 1981, Duncan Haldane realized that he could apply KT ideas “to the quantum mechanical 1D spin chain if he turned one of the spatial dimensions into time. Then the vortices of KT would become tunneling events between different topological states.” [19].

In 1986, neutron scattering was applied to a mixture CsNiCl which has magnetic 1D chains making it a quasi-1D compound and verified some of Haldane’s theories. He later discovered many interesting and unexpected new properties [17] which contributed to later advances in condensed matter physics and also had similarities to the 2D physics. Haldane was the youngest of the three winners (b 1951) and had studied under Philip Anderson. Examples of his 1D problems include chains of magnetic atoms, large spin Heisenberg anti-ferromagnet, chains of fermions versus bosons, 1D conductors (quantum wires and now carbon nanotubes), and 1D electron gas. His 1982 paper on spin chains showed topological properties due to “the collective action of the whole chain.” There are “topologically protected excitations that behave like Majorana fermions, which are their own antiparticle.” He has also been contributing to the understanding of the fractional quantum Hall effect (FQE). Advanced topological topics being used include: Chern Simons theory, O(3) non-linear sigma model, solitons, and instantons. And like the previous discussion, there are analogies of these solid state concepts in high energy physics. For example, Laughlin believes that the quark charges of ⅓ and ⅔ e may have an origin similar to that of the effectively fractional electron charges in the FQE.

References:

1. David Tong, **The Quantum Hall Effect**, TIFR Infosys Lectures, http://www.damtp.cam.ac.uk/user/tong/qhe.html, Jan 2016, 252 pages!
   [WAS ALSO 9 THEN DELETED, delete this note]
2. K. v. Klitzing, G. Dorda, and M. Pepper, “New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance,” Phys. Rev. Lett. 45, 494 – Published 11 August 1980.
3. J. Bellissard, et. al., “The Non-Commutative Geometry of the Quantum Hall Effect,” arXiv:cond-mat/9411052 14 Nov 1994, Feb 1, 2008. 78 pages.

4. Ad Lagendijk, eg. al., “Fifty years of Anderson Localization,” Physics Today, August 2009.
5. Steven M. Girvin, “The Quantum Hall Effect: Novel Excitations and Broken Symmetries, arXiv:cond-mat/9907002, 1998. 124 pages.
6. Dave Peterson, “Geometry in Modern Physics,” July, 2016, 20 pages.
7. S. Sugawa, et. al., “Observation of non-Abelian Yang Monopole: From New Chern Numbers to a Topological Transaction,” 2016 ArXiv 1610.06228 [NIST].
8. TKNN D.J. Thouless, M. Kohmoto, M.P. Nightingale, M.den Nijs, “Quantized Hall conductance in a Two-Dimensional Periodic Potential,” Phys. Rev. Lett, 49,#6, 9 August 1982, p. 405.
9. Wout Neutkens, “A historical perspective on the quantum hall effect,” 2013, Master at ITFA Amsterdam (supervisor AMM Pruisken).
10. Nii Amoo Dodoo-Amoo, “The nature of critical phenomena of the scaling theory of quantum Hall transitions in two-dimensional electron systems,” 2013, The University of Leeds, PhD Thesis. 245 pages.
11. R. B. Laughllin, “Quantized Hall conductivity in two dimensions,” Phys. Rev. B, 15 May 1981, pp. 5632-5633. More on the “Laughlin Gauge Argument” is found in [13].
12. Robert B. Laughlin, Fractional Quantization, Nobel Lecture: http://www.nobelprize.org/nobel_prizes/physics/laureates/1998/laughlin-lecture.html
13. J.M. Kosterlitz and D.J. Thouless, “Long range order and metastability in two dimensional solids and superfluids, J. Phys. C, Vol 5, 1972, p L124-L126.
14. Henrik Jeldtoft Jensen, “The Kosterlitz-Thouless Transition,” Course by Leonid Levitov, www.mit.edu/~levitov/8.334/notes/XYnotes1.pdf
15. F. D. M. Haldane, “Nonlinear Field Theory of Large-Spin Heisenberg Antiferromagnets: Semiclassically Quantized Solitons of the One-Dimensional Easy-Axis Néel State,” Phys. Rev. Lett. 50, 1153 – Published 11 April 1983 (APS free reading).
16. Masud Chaichian, Hugo Perez Rojas, Anca Tureanu, Basic Concepts in Physics from the cosmos to quarks, 2014, Springer.
17. APS Nobel Focus: “Topological Phases,” “Current for a Small Charge,” and “Topologically Speaking” http://physics.aps.org/topological-phases (Nov. 2016). [American Physical Society].
18. Foundational theories in topological physics garner Nobel Prize, Physics Today, December, 2016, Vol. 69 No 12. pgs 14-17 Search and Discovery. [The laureates bucked conventional wisdom]
19. F. Duncan M. Haldane - **Nobel Lecture “**Topological Quantum Matter,” Dec. 8, 2016
20. Joseph Avron, et al., “A Topological look at the quantum Hall effect.” Physics Today, August 2003, pp 38-42
21. APS Focus: Landmarks-- Accidental Discovery leads to calibration standard, May 15,2015, Physics.
22. Shou-cheng Zhang, Viewpoint: “Topological states of quantum matter,” physics.aps.org/articles/v16, 2008
23. Barrett O’Neill, Elementary Differential Geometry, 2nd Ed., Academic Press, 2006, 1997, 1966.

[This paper was written using the gmail Google Drive as a “Google Doc” with LaTex formulas for the Boulder Cosmology club at BPL].

davepeterson137@gmail.com

SPINORS

DAVID L. PETERSON

Abstract. Roger Penrose [1] intuitively defined a spinor as an object which turns into its negative after a complete $2\pi = 360^o$ rotation; and the action of rotation on a spinor is always double-valued. Here we consider two cases of spinor maps of Lie (continuous) groups: $SU(2) \to SO(3)$ for rotations and $SL(2, C) \to SO(1, 3)$ for Lorentz transformations (metric + - - -). Both of these maps are 2-to-1. Similar to the way a rotation matrix rotates a vector $\vec{x'} = R(\theta)\vec{x}$, two-component complex spinors transform like $\xi' = u\xi$, $u \in SU(2)$. The elements of SU(2) are 2×2 matrices containing half-angles, $e^{-i\theta/2}$, and that is imparted to spinor rotations. SU(2) spinors were intended to represent classical rotations but then also found application in Pauli's electromagnetic equation for quantum mechanics. An interpretation of the Dirac bi-spinor (superposition of two Weyl chiral L and R spinors) is a rapid zig-zag motion back and forth at light speed preserving the net handedness of rotation.

1. Introduction

Physicists typically associate the word "spinor" with the 4-component complex column vector,$\Psi(\vec{x}, t)$, used in Dirac relativistic quantum theory or the 2-component column vector used in non-relativistic Pauli spin algebra. The values contained in the spinor indicate the relative weights of the various spins (e.g., how much spin-up compared to how much spin-down). The name "spinor" relates to its first use by Klein for the classical spinning top in 1897. This was followed by its use in geometry by Èlie Cartan in 1913, Wolfgang Pauli's spin matrices in 1927 and Dirac's relativistic electron spin in 1928. Spinors are essential for understanding particles called "fermions" with half-integral spin requiring that particle revolutions have to "go twice around" to return to their initial state. And, understanding this goes with the idea of "double-cover" of groups.

Dirac 4-"spinors fully incorporate special relativity including Lorentz group of rotations and boosts" – they are built into the overall formalism [3]. The weighted strengths of the types of spin in Dirac spinors can depend on the degree of Lorentz boosting as functions of E, p, and m. Even general relativity theory finds use for spinors [4].

Throughout this note, we use matrices to "represent" elements of continuous groups. That means that for any two elements a and b of group G, matrix representations D(a) and D(b) must obey $D(a)D(b) = D(ab)$. A starting point for the idea of "double cover"

email: davepeterson137@gmail.com. January 06, 2018, update to 06/13/18.

is the continuous complex group SU(2) having a "2-fold cover" for for the three-space rotation group SO(3) [$S =$ "special," $U =$ unitary, and $O =$ "orthogonal"]. That means that every 3×3 rotation matrix of SO(3) maps to two different element of the 2×2 complex matrices of SU(2): group elements u and -u $\in SU(2)$. That is, there is a 2:1 homomorphism $SU(2) \rightarrow SO(3)$. We can say that SU(2) serves as a representation of SO(3), $g \rightarrow D(g)$, which itself can be represented by matrices.

The SU(2) group is isomorphic to Hamilton's quaternions (H) which possess three complex numbers "i, j, k" and was one of the first examples of "hypercomplex" number systems. These, quaternions, in turn, are a sub-algebra of "bi-quaternions" (complex quaternions, $C \otimes H$) which are isomorphic to the Pauli algebra $P = C\ell(3,0)R$ standing for three roots of plus one where $C\ell$ means "Clifford" algebra and "R" means "over the reals"]. Using the more familiar Pauli sigma matrices is more convenient and useful than just quaternions [themselves labeled as $H = C\ell(0,2)$ for having two imaginary roots of minus one — and then $e_1 e_2 = e_3$ catches the third imaginary]. The generators of the Lie algebra $su(2)$ are the i, j, k quaternions – but the Pauli matrices also suffice.

When considering 4-dimensional Minkowski space, relativistic Lorentz transformations generalize the role of rotations. For a metric of $(+ - --)$ we have $SO^+(1,3)$ – the orthochronous 4×4 Lorentz transformations $\Lambda^\mu{}_\nu$ preserving orientation and time direction $\Lambda^o{}_o \geq +1$. Its complex two-fold covering group is the special linear group $SL(2,C)$, and $SL(2,C) \rightarrow SO^+(1,3)$ is called the "spinor map." The group SL(2,C) is the set of all 2×2 complex matrices with unit determinant $\{M_2(C), det\, M = 1\}$ –not necessarily unitary.

2. Double Cover and Half Angles

A spinor is a mathematical object that turns into its negative when space is rotated through a complete turn, $S(\theta + 2\pi) = -S(\theta)$. Double cover means that it takes a 4π rotation to achieve what would classically be a 2π turn. A few simple examples of double cover includes:

<u>Möbius band:</u> The easiest example to visualize is the twisted "Möbius band" with up-arrows printed on the band. When rotated 360 degrees, the up arrow is seen as a down arrow, and return to origin takes 720^o –two rotations. If one placed a ball on a Möbius strip and made it rotate to the right along the strip, its initial spin would be "up". After moving through 2π, the ball will be seen as rotating down; and after 4π radians it will be up again.

<u>Euler:</u> Consider the square root of the Euler form $e^{i\theta}$, $\sqrt{e^{i\theta}} = e^{i\theta/2}$ (call "Root") which becomes negative after rotating angle theta through 2π radians ($e^{i\pi} = -1$). Notice that the mapping from Euler to "Root" is 2:1– double cover.

Two circles visual picture: Rotate one circle around a fixed circle of the same radius with a no-slip condition at the circumference point of contact. With the one set circle A centered at the origin, O_A, initially place the moving circle B to its right on the x-axis and mark the initial contact point with red ink. A rotating line segment L from the origin joins centers of both circles through the moving point of contact and advances counter clock-wise (ccw) at angle θ with circle B also rotating ccw. In the moving circle, the red mark will move downwards at angle θ with respect to line L. But rotation of circle B is with respect to a horizontal line H through its center O_B parallel to the x-axis, and the angle between H to L is also theta. So, the moving circle rotates twice as fast as the radial line L. When circle B has a full 2π rotation, it lies just to the left of A at angle $\theta = \pi$. This double-fast rotating is also clear by visual sketches. This concept is similar to Weyl's picture of a cone along a z-axis with vertex at 0 with another identical adjacent cone rotating about it at two rotations to one.

3. 3-space SO(3) and SU(2):

The rotation group is often introduced using three "Euler angles:" Rotate about a z-axis by angle ϕ, then rotate the z-axis to a new z-axis z' through an angle θ and finally rotate about the new z'-axis by angle ψ (some references instead use angles α, β, γ) [26]. Each rotation is labeled by an element g of the SO(3) 3d rotation group resulting in a rotated element $g = g(\psi)g(\theta)g(\phi)$ for net action $\vec{x'} = g\,\vec{x}$. In 1775, Leonard Euler showed "any displacement of a rigid body such that a point on the rigid body remains fixed, is equivalent to a single rotation about some axis that runs through the fixed point [8]." So the rotations through angles $\{\phi, \theta, \psi\}$ can be duplicated using just one rotation, ψ about some axis specified by direction cosines, $\hat{n}$. Alternatively, one can perform calculations just using hypercomplex quaternions (often preferred in industry and computer graphics). Calculations can be performed in 3d space or using equivalent SU(2) 2×2 complex matrices. Discussions of spinors occur with respect to complex matrices such as these.

The group SO(3) is a continuous group called a "Lie group" which may also be described by first specifying its "Lie algebra" $\mathfrak{so}(3)$ in terms of its basis matrices and then exponentiating to get the elements of SO(3). The Lie algebra is essentially the tangent space of the group near the identity (the 3x3 matrix I with all ones on the diagonal). So, the matrix shown below, $R_z(\theta) = e^{\theta E_3}$, has differential $dR_z \simeq d(I + \theta E_3) = E_3 d\theta$ for a tangent "vector" near I.

For infinitesimal rotations about the x-axis, y-axis and z-axis, a basis of the Lie algebra is a set of infinitesimal generators [9]:

$$E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix},\ E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix},\ E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1}$$

And as example, $R_z(\theta) = exp(\theta E_3) = \begin{pmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 0 \end{pmatrix}$ for an SO(3) rotation about the z-axis (like Euler angle ϕ above). These bases are related by commutators: $[E_i, E_j] = \sum_k \epsilon_{ijk} E_k$ (e.g., $E_1E_2 - E_2E_1 = E_3$).

Just as a 3-space vector can be rotated into another vector of the same length, $\vec{x'} = R(\theta)\vec{x}$, a complex space spinor can be transformed by an element of $SU(2)$ as $s' = u\ s$ where u is a 2×2 complex matrix $\in SU(2)$. We express that using Pauli sigma's:

The Pauli 'sigma' matrices, P, are fairly standardized and most often presented by:

$$\sigma^0 = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .\ \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_z = \sigma^3 = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

Pauli used a convention of representing electron spin with respect to a z-axis, so his σ_z is diagonal with elements $+1$ for spin up and -1 for spin down. If the elements of the 2×2 matrix σ_x are real, then the mathematics of spin forces the elements of σ_y to be imaginary. The three generating σ_i's, $i = 1, 2, 3$, satisfy the anticommutation requirement $\{\sigma_i, \sigma_j\} = 2\delta_{i\,j}\, I_2$. And commutation relations are given by $[\sigma_j, \sigma_k] = 2i\epsilon_{jk\ell}\sigma_\ell$ [9]. For example, $\sigma_1\sigma_2 - \sigma_2\sigma_1 = 2i\sigma_3$.

In quantum mechanics, Pauli matrices can be operators operating on spinors, e.g., for spinor $\xi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, $\quad \hat{\sigma_x} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$.

A 3-space vector $\vec{x} = (x_1, x_2, x_3)$ can be associated with a 2×2 complex Hermitian matrix, X, as follows:

$$X = \vec{x} \cdot \vec{\sigma} = x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3 = \begin{pmatrix} x_3 & x_1 - ix_2 \\ x_1 + ix_2 & -x_3 \end{pmatrix} \tag{3}$$

{Later, for relativity, including time or x^o would add an $x_oI = ctI$ to the diagonal of the matrix [called the $(+1, \vec{\sigma})$ basis – see Eqn.(8)]; and in relativity, $x^0 = ct$.} The 3-space matrix $\vec{\sigma} \cdot \vec{x}$ is called the "Pauli vector" and is an element of the the Clifford algebra generated by the Pauli matrices. Again, the basis of X are technically quaternions which may be re-expressed in terms of Pauli matrices. The matrix u may be expressed in terms of Euler angles: the angles θ and ϕ specify an $\hat{n}$ axis, and the rotation angle is now labeled as angle ψ [15].

As an example of the above, a 3d rotation of a vector (x,y,z) through some angle θ about the z-axis results in a new vector $\vec{x'} = R_z\vec{x} = (x', y', z' = z)$ resulting in
$$x' = xcos(\theta) - ysin(\theta), \quad and \;\; y' = xsin(\theta) + ycos(\theta).$$

This result can also be achieved by operating on the complex matrix X by $X' = u\,X\,u^\dagger$ where unitary matrix $u \in SU(2)$. "Dagger" † means "complex conjugate transpose," and the X transformation is an example of a "similarity transformation" [or inner automorphism or "conjugation: $x \mapsto gxg^{-1}$]. And In this case, let

$$(4) \qquad u = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{+i\theta/2} \end{pmatrix}, \quad with\, general\, form \;\; U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \in SU(2).$$

where $det(U) = |a|^2 + |b|^2 = 1$. [Note that if $U = a\sigma_z + \beta\sigma_x + c\sigma_y$, then $a = a_1 + ia_2$ obeys $a^* = -a$ so that $a = ia_2$ (pure imaginary)]. The form above says what U's and u's can be and excludes the matrix forms such as X, and Q and N shown below. The matrix u effectively uses the z quaternion (shown below in Eqn.(32)) times theta times electron spin ($\frac{1}{2}$, and we can assume $\hbar = c = 1$ as chosen units). Then that $\theta e_3/2$ gets exponentiated for the element of the SU(2) Lie group.
Calculating the product of three matrices $u\,X\,u^\dagger$ by hand indeed yields the correct expression for the rotation from $\vec{x}$ to $\vec{x'}$ and shows the consistency of using half-angles.

In texts [4], they are often introduced by the relation between rotations and reflections through planes. "A rotation through an angle θ about a given axis may be visualized as the consequence of successive reflections in two planes that meet along the axis at the angle $\theta/2$." [4]. The general form with a,b terms above follows from the unitary requirement that $U^\dagger = U^{-1}$. If we were to use the negative of the unitary transformation, we would also get the same degree of rotation: $(-u)v(-u^\dagger) = uvu^\dagger$. So two elements of SU(2) map to the same rotation element of SO(3) – double covering.

The example above for the 2×2 complex matrix $u \in SU(2)$ is also equal to $exp[-i\sigma_z\theta/2] = \Sigma(-i\theta/2)^n{\sigma_z}^n/n!$ from $n = 0$ to ∞. To perform this series expansion, we make use of the fact that the Pauli matrix ${\sigma_z}^2 = I$ and $e^{i\theta/2} = cos(\theta/2) + i\sin(\theta/2)$. The functions sine and cosine also possess odd and even infinite series expansions. Again, since the base z quaternion $k = -i\sigma_z$, the $u = exp[-i\sigma_z\theta/2] = exp[k\theta/2]$ is really a quaternion rotation (the "natural" language for SU(2)).

A standard full example of the general complex SU(2) matrix Q with elements "a" and "b" above, equation (4), incorporates the usual convention of "Euler" angles, ϕ, θ, ψ into something resembling "Cayley-Klein" parameters:

$$(5) \qquad a = cos(\theta/2)exp[(\psi + \phi)i/2], \qquad b = sin(\theta/2)exp[(\psi - \phi)i/2].$$

where , $0 \le \phi \le 2\pi$, $0 \le \theta \le \pi$, $0 \le \psi \le 4\pi$ (!).
And allowing angle ψ to go around twice is an indication of the double cover. Spinors are objects that transform under SU(2) elements like u or U (a lowest dimension spinor

representation). The Euler angle rotation scheme is separate from the quaternion (i, j, k) rotations.

There is an isomorphism between $SO(3)$ and $SU(2)$ rotations:
$R = exp(i\vec{J} \cdot \vec{n}\theta) \leftrightarrow exp(i\sigma \cdot n\theta/2)$ where the "n" unit direction in SU(2) has a form like the X-matrix [eqn.(3)], $N = \vec{\sigma} \cdot \vec{n} = \begin{pmatrix} n_3 & n_1 - in_2 \\ n_1 + in_2 & -n_3 \end{pmatrix}$. In place of $n\theta$, some texts just use $\vec{\theta}$ (which is often confusing). An effective SU(2) rotation is $R_n(\theta) = Icos(\theta/2) - i\vec{\sigma} \cdot \vec{n}sin(\theta/2)$.

Topology of SU(2) and SO(3) :
If all closed loop paths on the surface of a sphere can be contracted to a point (path deformation retract), the sphere is said to be "simply connected" and has a "fundamental group" $\pi_1(S^2) = 0$ [$\pi_1(S^3) = 0$ too]. Most arbitrary curves or paths around a circle S^1 cannot be contracted and may end up going around it an integer number of times, so $\pi_1(S^1) = Z$ (the group of integers). Are the groups SU(2) and SO(3) simply connected? Yes for SU(2) but no for SO(3).

A 3-sphere can be expressed by $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$; or, in complex space, it can be written as $\{z_1, z_2 \in C : |z_1|^2 + |z_2|^2 = 1\}$. It is a 3d object (for example like angles, ϕ, θ, ψ) embedded in 4d space– but we can only perceive 3d space. Equation (4) above showed a general form for the elements of SU(2). In that form, we can substitute $a = z_1, and {-b^*} = z_2$ for an equivalent element $u \in SU(2)$:

$$\begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} = \begin{pmatrix} z_1 & -{z_2}^* \\ z_2 & {z_1}^* \end{pmatrix} \rightarrow \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in S^3 \tag{6}$$

That is, simply map the matrix u to just its first column. Since S^3 is simply connected, so is $SU(2)$ and $\pi_1(SU(2)) = 0$.

The group of 3-space rotations $SO(3) \subset O(3)$ which is a larger group having $det\, R = \pm 1$ (versus "special" with det = +1). [$det\ R = -1$ is called improper and includes "space inversion," $I_s = -I$ where $I_s\vec{x} = -\vec{x}$ (obviously no proper SO(3) rotation could do that)].

A visual example: Let any rotation of a vector $\vec{x}$ about an axial direction $\hat{n}$ correspond to a point, P, inside a 3d ball (3-ball or 3-"disk") such that the length represents an angle $OP = \theta,\ 0 \leq \theta \leq \pi$ stands for the angle of rotation. That means that the radius of the ball is now $r = \pi$ and has a surface sphere of radius π, [$S^2 = \partial B^3$, B = "ball"]. A 180^o rotation has $\vec{x}' = R\vec{x}$ where $R' = OP'\ at\ \theta = -\pi$, where P' is opposite to P on the sphere (antipodal). These points are identified $P = P'$ because they yield the same rotation. This is called "real projective 2-space" with projection $P' \rightarrow P$. In particular, north and south poles of the sphere are identified, $S = N$.

Now, like the usual sphere, most closed curves originating at $P = N$ can be retracted to a point (so simply connected). But, an open curve between N and S (which is then closed since $N = S$) cannot be deformed to a point because it's length is mostly constrained on the surface S^2. Therefore, $SO(3)$ is not simply connected.

SU(2) Spinors:

Two component spinors $\xi = \xi^i = \begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix}$ transform by $\xi' = u\xi$ meaning $\xi'^i = u_{ij}\xi^j$ with $u \in SU(2)$. These are called "contravariant spinors of rank 1." To define a scalar product, form a "covariant" rank 1 spinor η_i transforming like: $\eta' = \eta u^{-1} = \eta u^\dagger$ [14]. Then $\eta\xi = \eta'\xi' = \eta_i\xi^i$. And $\eta_i{}' = u^\dagger{}_{ji}\eta_j = \eta^*{}_{ij}\eta_j$. This is also the way that ξ^* transforms: $\xi'^{*j} = u^*_{ij}\xi{*}^j$ (u* is called the conjugate representation). So, $\xi^{*i} \equiv \xi_i$. $u^* = SuS^{-1}$ – a similarity transformation where $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ Then $\bar{\xi} \equiv S^{-1}\xi^* = \begin{pmatrix} -\xi_2 \\ \xi_1 \end{pmatrix}$ (single application of S for spinors). One can form an "outer product" $\zeta_i^j = \xi^j\xi_i$ that can be made to be "traceless" by forming $\hat{\zeta}_i^j = \zeta_i^j - \delta_i^j \sum_k \xi^k\xi_k/2$ [14]. It is claimed that $\hat{\zeta}_i^j$ is a 4-component tensor that can be identified with $X = \vec{\sigma} \cdot \vec{x}$.

Note that a spinor $\xi_\perp = \begin{pmatrix} -\xi_2{}^* \\ \xi_1{}^* \end{pmatrix}$ is orthogonal to ξ. That is,

$$\langle \xi_\perp, \xi \rangle = (\{\xi_\perp{}^\dagger = \xi_\perp{}^{*T}\})\,\xi = (-\xi_2, \xi_1) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = -\xi_2\xi_1 + \xi_1\xi_2 = 0.$$

Linear algebra [see Appendix] says that we can form a "transformation matrix" for a change of basis by columns of eigenvectors: $P = (\xi,\, \xi_\perp)$ so that a new matrix $A' = PAP^\dagger$ (called a similarity transformation). We put this back into the usual a,b labels form (modified to agree with a standard source, "Steane" [12]), and normalize it to a unitary transformation [so in eqn.(4), we let that $U_{old} \longrightarrow U^*$, $P = U_{old}{}^T$ and the Caley convention [eqn.(5)] gets conjugated]. This is ok; it is all still in SU(2) form.

Determinant $det(P) = |a|^2 + |b|^2$ and let $d = \sqrt{det(P)}$. Then, transformation matrix $V = P/d$:

(7)

$$\vec{s} = \begin{pmatrix} a \\ b \end{pmatrix} = de^{-i\psi/2} \begin{pmatrix} cos(\theta)e^{-i\phi/2} \\ sin(\theta)e^{\,i\phi/2} \end{pmatrix},\ S = V\sigma_z V^\dagger = \frac{1}{d^2} \begin{pmatrix} |a|^2 - |b|^2 & 2ab^* \\ 2ba* & |b|^2 - |a|^2 \end{pmatrix}.$$

This is in the context of "spinors as flagpole with flag rotating about the flagpole" where the length of the flagpole is $d^{\,2}$ and the flag angle is the last Euler rotation angle, ψ. The radial flagpole vector is $\vec{r} = (r_x, r_y, r_z)$ with $|\vec{r}| = d^{\,2}$. Algebraic inversion of the terms in spinor $\vec{s}$ give $r_x = ab^* + ba^*$, $r_y = i(ab^* - ba^*)$, $r_z = |a|^2 - |b|^2$ which are now recognized as terms in the matrix S above. After normalization, $\vec{r} = \vec{n}$ as a unit cosine directional vector for rotation. That means that matrix S is a spin matrix that has a direction $\vec{n}$ (when written out in the form for matrix X in eqn.(3)). This may also be expressed as

$\vec{n} = \langle s|\vec{\sigma}|s\rangle/d^2 = s^\dagger\vec{\sigma}s/d^2$.

$S\vec{s} = +1\vec{s}$ and $S\vec{s} = -1\vec{s}$ (or, in previous notation, $S\xi = +1\xi$ and $S\xi_\perp = -1\xi_\perp$). Therefore, it has been shown that "Every spinor is the eigenvector with eigenvalue ± 1, of a 2×2 traceless Hermitian matrix" $\{S\} = \vec{n}\cdot\vec{\sigma}$ [12]. And "the direction associated with the matrix will agree with the flagpole direction of the spinor!"

4. **Lorentz Group and** $SL(2,\mathbb{C})$

The special linear group $SL(2,\mathbb{C})$ is the set of all 2×2 complex matrices with unit determinant, $SL(2,\mathbb{C}) = \{g \in M_2(\mathbb{C})|det(g) = 1\}$. Unlike the Lie group $SU(2)$, it is not required to be unitary [i.e., where $U^{-1} = U^\dagger = \bar{U}^T$]. The mapping of $SL(2,C) \to SO^+(1,3)$ is called its "spinor map." Typical $sl(2,C)$ generators for the $M_2(\mathbb{C})$ matrices are the matrices E, F, H in eqn.(10) below.

"The Lorentz group is a 6-dimensional Lie group of linear isometries of the Minkowski space" and can map to the condensed matrix group $SL(2,C)$. In mathematical physics, the Lorentz group is the set of all relativistic Lorentz transformations with any Lorentz transformation being the product of a pure rotation and a pure boost, $\Lambda = RB$. A matrix representation $\Lambda \to D(\Lambda)$ is also a representation of the rotation group O(3) — not SO(3) because boosts are not unitary.

A 4-space vector $\vec{x} = (x_o,\ x_1, x_2, x_3)$ can be associated with a 2×2 complex Hermitian matrix, Q, as follows:

$$Q = \vec{x}\cdot\vec{\sigma} = x_o\sigma_0 + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3 = \begin{pmatrix} x_o + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_o - x_3 \end{pmatrix}, \tag{8}$$

This is like the previous 3-space rotation of vector X of eqn.(3), but now including time or x^o by adding an x_oI to the diagonal of the matrix [this is called the $(+1,\vec{\sigma})$ basis; and in relativity, $x^0 = ct$]. For SL(2,C), matrix Q has determinant $det(Q) = (x^0)^2 - x^2 - y^2 - z^2$ recognized as a relativistic invariant length which is preserved under the action of elements of SL(2,C) [with the pre-selected convention of metric signature $(+,\ -\ -\ -)$]
For matrix Q, we transform as $Q' = gQg^\dagger$ where $g \in SL(2,C)$ and $g = g_\mu\sigma^\mu$.

A general condensed form for the g elements is $g(\rho,\vec{\theta}) = exp[(\rho/2 - i\hat{n}\theta/2)\cdot\vec{\sigma}]$.
The symbol ρ is boost "rapidity" and is given by $\rho = \pm tanh^{-1}\beta$ where $\beta = v/c$. Since $sinh^2(\rho) - cosh^2(\rho) = 1$, we have $cosh(\rho) = \gamma = \sqrt{1/[1-\beta^2]} = E/m$.

$$g(\rho,\vec{0}) = \begin{pmatrix} e^{-\rho/2} & 0 \\ 0 & e^{+\rho/2} \end{pmatrix} = cosh(\rho/2)I - sinh(\rho/2)\sigma_z. \tag{9}$$

This boost in the z-direction doesn't make use of the imaginary i .

Technically, one should say that the group of Lorentz transformations $L = L_+^\uparrow = T(\Lambda) \cong SL(2,C)/Z_2$ [5] (1-1, onto, double cover removed) as its complex representation.

A basis of infinitesimal generators for its Lie algebra $sl(2,C)$ is:

$$(10) \qquad E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\ H = \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad M = \begin{pmatrix} \gamma & \alpha \\ \beta & -\gamma \end{pmatrix}$$

The Lie Group elements are formed by exponentiating the matrix representing the Lie algebra: $M = \alpha E + \beta F + \gamma H, \quad \{\alpha, \beta, \gamma\} \in \mathbb{C}$.
Since $\sigma_x = E+F$ and $\sigma_y = i(F-E)$, this also contains the Pauli matrices (or quaternions) needed for SU(2). So, $SU(2) \subset SL(2,C)$ [e.g., $sl(2,C) = su(2) + i\, su(2)$]. The form of M contains the general "a,b" form for su(2) rotation matrices "U" [eqn. (4)] by setting $\gamma = i\, Im(a)$, $\alpha = b$, $\beta = -b^*$. Alternatively, a 6-dimensional basis with 2×2 matrices is $J^i = \sigma^i/2\; and\; K^i = i\sigma^/2$ to include boosts.

The special relativistic Lorentz group, L, consists of transformations, $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ with $\Lambda^\mu{}_\nu = \Lambda = RB$ which are 4×4 matrices whose six infinitesimal base generators have elements of $-i \times \{-1, 0, +1\}$ that perform two functions: rotating spatial 3-vectors and boosting 4-vectors by a new speed: There are three anti-symmetric rotation matrices J^1, J^2, J^3 and three symmetric boosts K^1, K^2, K^3. Once these are known, exponentiation give the general transformation: e.g., $U(R_\theta) = e^{iJ\cdot\theta}$. Note that Weinberg's notation for generators [5] is $J = \{J^{23}, J^{31}, J^{12}\}$ and $K = \{J^{10}, J^{20}, J^{30}\}$. These 4×4 matrices are commonly presented in texts on General Relativity or Quantum Field Theory.

Homomorphism of SL(2,C) on the Group L:
Carmeli's [15] mapping from g's to Lorentz transformations are: $\Lambda^\alpha{}_\beta = ({}^1/{}_2) Tr(\sigma^\alpha g \sigma^\beta g^\dagger)$.
For example, $\Lambda^o{}_o = |g_o|^2 + \Sigma |g_k|^2$. Define a normalized eigenvector, $\vec{v_o} : (v_o)^k = \Lambda^k{}_o / \sqrt{\Sigma(\Lambda^\ell{}_o)^2}$ and $\vec{v_1} = \Lambda^o{}_o + 1$. Then element $g_k = \pm[function\; of\; v's]$ indicating a 2:1 mapping.

For a complexified Lie algebra with new component bases, form two new matrices: $A = (J + iK)/2$ and $B = (J - iK)/2$ – these satisfy the commutation relations of su(2) and so(1,3). Using these, an SO(1,3) representation can be classified as a $SU(2) \otimes SU(2)$ representation.

In the previous discussion, a rank 1 spinor transforms as $\vec{s'} = \Lambda \vec{s}$ under a change in inertial frame. The outer product $ss^\dagger$ is a 2x2 matrix 2nd rank spinor transforming as $ss' \longrightarrow \Lambda ss^\dagger \Lambda^\dagger$. The previous matrix Q (eqn.(8)) is a 2nd rank spinor, $Q' = \Lambda Q \Lambda^\dagger$. Q and $ss^\dagger$ can be equated to create an associated 4-vector [12]. Part of this procedure is the same as the previous association of the spin-matrix S (eqn.(7)) with the components of the radial flagpole vector, $\vec{r}$.

$$ss^\dagger = \begin{pmatrix} a \\ b \end{pmatrix} (a^*, b^*) = \begin{pmatrix} |a|^2 & ab^* \\ ba^* & |b|^2 \end{pmatrix} = Q = \vec{x^\mu} \cdot \vec{\sigma^\mu} = \begin{pmatrix} t+z & x-iy \\ x+iy & t-z \end{pmatrix}, \tag{11}$$

where $\sigma^\mu = (I, \sigma^i)$. Solving this equation for t,x,y,z as a 4-vector gives:

$$\begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (|a|^2 + |b|^2)/2 \\ (ab^* + ba^*)/2 \\ i(ab^* - ba^*)/2 \\ (|a|^2 - |b|^2)/2 \end{pmatrix} = \langle s|\sigma^\mu|s\rangle/2. \tag{12}$$

Now the determinant $det(ss^\dagger) = 0$ suggesting that the 4-vector x^μ is null. And this suggests that one can obtain a null 4-vector V from a right-handed contra-spinor v as $V^\mu = v^\dagger \sigma^\mu v$.

5. Representations

There are many different conventions for Lie Group representations.
First consider simple spin one-half and spin zero:

For spin $^1/_2$, the "left" Weyl spinor (e.g., for neutrinos) is said to be in the Lorentz representation $\ell = (\frac{1}{2}, 0)$ while the "right" Weyl spinor is in the Lorentz representation $r = (0, \frac{1}{2})$ (e.g., for anti-neutrinos). A general 4d Dirac fermion representation is $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2}) = \ell \oplus r$ – also called a bispinor representation.

The "Lorentz representation" for scalars (e.g., Higgs boson) with spin 0 is $(0, 0)$.
Photons with circular polarizations pointed backwards or forwards may be in the Lorentz transformation representation $(1, 0) \oplus (0, 1)$ of SL(2,C). Specifically, that description "can provide a 6-component spinor equivalent to the EM field tensor", F.
One could add total spin 1 for "vector gauge fields" (γ, Z, $W^\pm$, *gluons*) with Lorentz representation $(\frac{1}{2}, \frac{1}{2})$ [with respect to the group SO(4) !].

There are fairly standard short notations in particle physics for discussing combinations of particle "flavors." The common Lie groups are SU(2) with 2×2 matrices and SU(3) with 3×3 matrices. For example, we know that the number of elements in a square matrix or tensor, T, n^2 can be decomposed into symmetric and anti-symmetric (or skew) elements according to [16]:

$$S^{ij} = \tfrac{1}{2}(T^{ij} + T^{ji}),\ A^{ij} = \tfrac{1}{2}(T^{ij} - T^{ji}),\ n^2 = \tfrac{1}{2}(n^2 + n + n^2 - n) = \tfrac{1}{2}n(n+1) + \tfrac{1}{2}n(n-1). \tag{13}$$

So, for n = 2, $\mathbf{2} \otimes \mathbf{2} = \mathbf{3} \oplus \mathbf{1}$ for symmetric plus skew parts (as mentioned in the introduction above). $\otimes$ refers to tensor products and $\oplus$ refers to direct sums. A particular example [10] is the addition of two spin $\frac{1}{2}$ particles with spin states $\uparrow$ & $\downarrow$ where the symmetric total spin one states have spin projections $M_s = 1, 0,$ & -1 (a triplet) while total spin zero is a single anti-symmetric state:

$|S = 0, M_s = 0\rangle = \sqrt{\frac{1}{2}}\,(\uparrow\downarrow - \downarrow\uparrow)$.

And for n = 3, $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \mathbf{3}$ or sometimes written as $\mathbf{6} \oplus \bar{\mathbf{3}}$, where the bar emphasizes anti-symmetry under the exchange of two given particles. Some use the over-bar for anti-particles so that a **3** representation for u,d, and s quark flavors would have $\bar{3} = \{\bar{u},\ \bar{d},\ \bar{s}\}$. If we are instead working with QCD colors, then $3 = \{r,\ g,\ b\}$ and $\bar{3} = \{\bar{r},\ \bar{g},\ \bar{b}\}$ anti-colors.

The group SU(3) has 3×3 unitary, unimodular matrices whose 8 independent generators are usually chosen to be the 8 Gell-Mann matrices, λ_i. This is similar to the use of the Pauli matrices $\{\sigma_x,\ \sigma_y,\ \sigma_z\}$ for SU(2).

One often wishes to treat the trace of a tensor separately so that a tensor product of two vectors is decomposed as the addition of a traceless-symmetric part + the trace of the anti-symmetric part. Then the above example for n = 2 becomes $\mathbf{2} \otimes \mathbf{2} = \mathbf{2} \oplus \mathbf{1} \oplus \mathbf{1}$
And for n = 3, $\mathbf{3} \otimes \mathbf{3} = \mathbf{5} \oplus \mathbf{1} \oplus \mathbf{3}$ More complicated tensor products are often treated using something called the "Young Tableaux."

We can also combine three quark flavors together; e.g., The total number of states formed from $\{u,\ d\ ,s\ \}$ is $3 \times 3 \times 3 = 27$. These can be decomposed in a number of ways. Overall, the matter baryons are symbolized by [1]:

(14) $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = (\mathbf{3} \otimes \mathbf{3}) \otimes \mathbf{3} = \mathbf{27} = (\mathbf{6} \oplus \mathbf{3}) \otimes \mathbf{3} = (\mathbf{6} \otimes \mathbf{3}) \oplus (\mathbf{3} \otimes \mathbf{3}) = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}$

(including Gell-Mann octets – the "Eightfold Way.")

"Block Diagonal Matrices" (BDM) are incorporated in another mathematical scheme showing representations using square matrices for which non-diagonal elements are all zero. The objects along the diagonal can be other square matrices of various sizes. A nice property of these BDMs is preservation of the general form of the BDMs under operations:

If the objects on the diagonal [**diag( )**] of matrix M are square matrices A, B, C, we say $M = diag(A, B, C)$ or $M = diag(A_{11}, A_{22}, A_{33})$.
Squaring matrix M preserves its shape: $M^2 = diag(A^2, B^2, C^2)$ [17]. This works for exponentiation e^M and for inverses too: $M^{-1} = diag(A^{-1}, B^{-1}, C^{-1})$.

It is also conveniently true that determinant $detA = detA_{11} \times detA_{22} \times \ldots \times detA_{nn}$; and trace is the sum: $trA = trA_{11} + trA_{22} + \ldots$

If the representation of a given operator is a block diagonal matrix, then it is called a **reducible** representation. Otherwise it is called an irreducible one. A reducible representation decomposes the vector space V it is acting on a direct sum $V = V_1 \oplus V_2 \oplus V_3 \oplus \ldots$

[1] Yes, $6 \otimes 3 = 10 \oplus 8$ can be shown using the Young Tableaux (e.g., Palash Pal, An Introductory Course of Particle Physics, CRC, 2015, HW p 267).

Each block acting on a subspace V_i. For all Lie groups, a representation which can be transformed to block diagonal form by means of a similarity transformation, i.e. $SUS^{-1} = diag(U(A), U(B))$, will be called reducible. "The algebra of the block matrix is frequently reduced to the algebra of the individual blocks." The fundamental representations for SU(n) are all irreducible, i.e. they can not be transformed to block diagonal form. For SU(2) the representations corresponding to spin $s = 1/2, 1, 3/2, 2, \ldots$ are irreducible. These are also labeled as $\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \ldots$ for the number of basis vectors or kets, $2j + 1$ in each subspace labeled by "j".

6. **Dirac Theory**

The most common key matrices representing Dirac Theory use the "Dirac-Pauli Standard" convention (DP-std) expressed as:

$$\alpha = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}, \ \beta = \gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \ \gamma^i = \gamma^0 \alpha^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}. \tag{15}$$

This "Dirac representation" for γ matrices [e.g., [11][21]] applies to high-energy massive particles physics (the usual HEP). Like other Clifford algebras, the four gamma matrices γ^μ generating the algebra must obey an anticommutation requirement $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} I_4$ where eta is the space-time metric and I_4 is a 4x4 unit matrix.

The Dirac equation is given by the Hamiltonian $H\psi = (\vec{\alpha} \cdot \vec{p} + \beta m)\psi = E\psi$ where α and β are 4×4 complex matrices invented by Paul Dirac in 1928.
Standard quantum physics operators include $\hat{E} = i\hbar\partial/\partial t$, $\hat{p} = -i\hbar\nabla$, $\partial_\mu = \partial/\partial x^\mu = \{\partial_o, +\partial_i\}$. Now, multiply the Dirac equation by β to give $(i\gamma^\mu\partial_\mu - m)\psi = 0$.
Paul Dirac's memorial stone in Westminster Abbey is inscribed with this momentous equation: $i\gamma^\mu\partial_\mu\psi = m\psi$.
In addition there is a matrix called gamma- 5: $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$ – an "exchange" matrix. This is important for discussing "chirality;" and the Dirac spinor in the Dirac representation has mixed chirality. There are "chiral projection operators" of the form $P_L = (1 - \gamma_5)/2$ and $P_R = (1 + \gamma_5)/2$ to project out the chiral handedness of a standard Dirac-Pauli spinor: $u_L = P_L u$ and $u_R = P_R u$. A simpler operator is used for Weyl spinors shown below. [For electroweak theory, the heavy boson W's only couple to L particles and R-handed antiparticles]. Old particle physics texts had an $x_4 = ict$ and used γ_4 for γ^0 – before the days of "FAREWELL ict " [4].

The above condensed form 2×2 matrices represent 4×4 matrices of complex numbers. The Dirac spinor (or "bi-spinor") ψ is a 4 component column matrix with the upper two entries representing the electron spin-up and spin-down and the bottom two entries represent the spin of the anti-electron or positron – sometimes stated as negative energy states.

As a first look at the 4-spinor of the Dirac equation, consider the case of a rest frame with no momentum [$v = 0$, $p = 0$, and then later we will modify this by applying a Lorentz boost, $\Lambda(p)$]. The equation having $\alpha \cdot p = 0$ becomes simply [11]:
(16)

$$H\psi = E\psi = \beta m\psi = \begin{pmatrix} mI & 0 \\ 0 & -mI \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \; so \; \psi = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

with eigenvalues $E = m, m, -m, -m$ and eigenvectors $u^1(0)$, $u^2(0)$, $u^3 = v^2(0)$, $u^4 = v^1(0)$. Free fermion wavefunctions are a product of a boosted fermion spinor times a plane wave: $\Psi(x^\mu) = u(p^\mu)e^{-ip \cdot x}$. An example of this for the rest frame is: $\psi^1 = u^1 \, e^{-imt}$ for the u^1 shown above. And spinors, of course, are not 4-vectors (they do not represent t,x,y,z).

The $E = -m < 0$ solutions are interpreted to represent antimatter. The common teaching about the positions in the standard Dirac spinor are: left-handed and spin-up, left-handed with spin down, right-handed and spin up, and right-handed with spin-down.

There is also a "charge conjugation" operator which flips signs of particle charges and changes a particle into an anti-particle: $\hat{C} = i\gamma^2\gamma^o$. When operating on electron spin up $u_1 \, e^{i(x \cdot p - Et)/\hbar}$ (a 1 in the upmost position), it produces $v_1 \, e^{-i(x \cdot p - Et)/\hbar}$ (with a 1 in the bottom position). That is why $u_4 = v_1$ and $u_3 = v_2$ as labels. Particle eigenspinors become anti-particle spinors.

The usual condensed 2x2 form of the general Dirac equation $i\gamma^\mu \partial_\mu \psi = m\psi$ can be written as:

$$Hu_{\,Dirac} = \Big[E \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} - m \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix} p^i \Big] \begin{pmatrix} u_A \\ u_B \end{pmatrix} = 0 \tag{17}$$

And, after expanding and negating some signs, this becomes

$$Hu = \begin{pmatrix} m & \vec{\sigma} \cdot \vec{p} \\ \vec{\sigma} \cdot \vec{p} & -m \end{pmatrix} \begin{pmatrix} u_A \\ u_B \end{pmatrix} = E \begin{pmatrix} u_A \\ u_B \end{pmatrix}, \tag{18}$$

With $m > 0$, this Dirac equation represents two coupled equations:

$$\vec{\sigma} \cdot \vec{p} \, u_B = (E - m)u_A, \quad and \quad \vec{\sigma} \cdot \vec{p} \, u_A = (E + m)u_B. \tag{19}$$

Consider the special case of a highly relativistic limit [27], $E \gg m$, $E \simeq p$, and choose to have momentum in the z direction accompanied with the diagonal sigma z. Then the equations become just $u_A = \sigma_z u_B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} u_B$ and $u_B = \sigma_z u_A$. For spinor solutions, successively set $u_A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $u_A = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $u_B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $u_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ for spinors u^1, u^2, u^3, u^4. Then the resulting spinors are:

$$u^1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\ u^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix},\quad u^3 = v^2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\quad u^4 = v^1 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}. \tag{20}$$

The additional 1's beyond the rest frame singles in the spinors shown in eqn.(16) materialize with the mathematically full Lorentz boosts from $v = 0 \rightarrow to\, v \rightarrow c$ where $p/E \rightarrow 1$. In-between these extreme cases of $p = 0$ and $p_z = E$ are spinor forms containing algebraic entries of p's and E's such as the Feynman spinor shown in equation(31) below. An interpretation of the multiple 1's in each spinor, in part, is that the energy $E \gg m$ is so high that the probability of antimatter becomes the same as that of matter [subject to conservation laws for net charge, energy/momentum, and angular momentum in possible interactions per Feynman diagrams].

Another examination of this $m = 0$ or $m \ll E$ of highly relativistic spinor/particle is to view a first order equation that is symmetric in space and time: the "Weyl" equation: $\sigma^\mu \partial_\mu \psi = 0$. For this special case of zero mass $m = 0$, Weyl noticed that the coupled Dirac equations [eqn.(19),$m = 0$] become decoupled. For decoupling, add and subtract these equations to get Weyl spinors:

$$\chi = u_B - u_A,\ and\, \phi = u_A + u_B; \quad so\, u_B \rightarrow \tfrac{1}{2}\,(\phi_R + \chi_L),\ u_A \rightarrow \tfrac{1}{2}\,(\phi_R - \chi_L) \tag{21}$$

Dirac standard spinors are mixtures of Weyl chiral spinors.

Mixing of χ_L and ϕ_R Weyl spinors into parts of the Dirac representation bispinor eqn.(21) requires an interpretation and be seen as a rapid back-and-forth "Zig-Zag" oscillation at light speed between an L rotation (left-hand thumb down) and an R rotation (right-hand thumb up) with a net rotation of fingers that is the same and preserved for both hands [28]. This picture is similar to the same L and R zig-zags of particles in a Higgs field that also preserve a net direction of electron spin for each spin type. In that case, the rapidity of vibration correlates with particle mass as the degree of coupling to the Higgs field. For the Dirac case, the vibration may be called "Zitterbewegung" ("jitter motion"– suggested by Schrödinger).

Or, for a simpler intuitive approach: The "square-root" of the massless Klein-Gordon equation results in both positive and negative energy cases, so there can be two solutions: spinors χ and ϕ. $\partial\psi/\partial t = \pm\vec{\sigma} \cdot \nabla\psi$.

We now have two equations for the two-component spinors χ and ϕ:

$$\hat{E}\chi = -\vec{\sigma}\cdot\vec{p}\,\chi,\ \ and\ \ \hat{E}\phi = +\vec{\sigma}\cdot\vec{p}\phi. \qquad (Weyl, 1929). \tag{22}$$

And, for massless particles, $E = |p|$ (where we use $c \equiv 1$ units).
Then $\vec{\sigma}\cdot(\vec{p}/|p|)\,\chi = -\chi$ and $\vec{\sigma}\cdot(\vec{p}/|p|)\,\phi = +\phi$.
The $-\chi$ case goes with negative "helicity," and the $+\phi$ goes with positive helicitiy — spin projected either against or along with the direction of momentum.

The negative helicity solution corresponds to the left-handed electron neutrino. The word "chirality" corresponds to "helicity" for massless particles moving with light speed. But, for particles with mass, helicity is not Lorentz invariant because a reference frame can be found exceeding the speed of the particle– and then helicity is effectively reversed. Chirality is a Lorentz invariant (with an unfortunate name that suggests helicity).

Some sources begin Dirac theory by defining the Dirac bi-spinor to be the stack of two Weyl spinors: $\Psi_D \equiv \chi_L \oplus \phi_R$ which transforms under a diagonal Λ chiral basis. The Dirac bi-spinor has 4 components, and the 4d representation of Λ can refer either to the standard representation or to the stacked Weyl representation. So, to specify a spinor, one must also state the representation: standard Dirac-Pauli or Weyl chiral. There is a transformation matrix A between the two:

$$\Psi_{DiracRep} = A\,\Psi_{WeylRep} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \phi_R \\ \chi_L \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} \phi_R + \chi_L \\ \phi_R - \chi_L \end{pmatrix}, \tag{23}$$

Weyl's equations were initially rejected by Wolfgang Pauli for violating parity before the awareness in 1956-1957 {Lee, Yang, Wu, Lederman,..} that nature really can violate parity in weak interactions, and particles can acceptably be left-handed or right-handed. With that awareness, the chiral gamma basis found increasing use against the standard Dirac gamma basis.

While the coupled Dirac equations (eqn.(18), $m > 0$) uses the standard Dirac-Pauli gamma basis (15), the uncoupled equations (22) prefer the use of a different set of gamma matrices called the "Weyl basis" or "chirality basis" (with some variation in conventions from text to text such as negating the γ^i matrix). In this representation, chirality gamma-5 is now block diagonal $-$, $+$.

$$\gamma^0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},\ \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix},\ \gamma^5 = \begin{pmatrix} -I & 0 \\ 0 & +I \end{pmatrix} \tag{24}$$

Unfortunately, there are a variety of conventions for expressing the chiral gamma matrices (not standard), and we might state them by a "top row sign" label order of {0, i, 5}. For reference, the "Standard Dirac-Pauli" convention for gammas [eqn.(15)] has all positive terms on their top rows for a label of $\{+,+,+\}$. A preferred set of chiral gamma matrices can be defined by ${\gamma_c}^0 = +{\gamma_D}^5$ and ${\gamma_c}^i = {\gamma_D}^i$, ${\gamma_c}^5 = -{\gamma_D}^0$ with top row signs $\{+,+,-\}$, e.g., [10][21][23]. Obviously, a sign change in either γ^0 *or* γ^i will produce a sign change in $\gamma_5 = \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. So, for an alternate choice of ${\gamma_c}^i = -{\gamma_D}^i$ and ${\gamma_c}^5 = +{\gamma_D}^0$ we get $\{0, i, 5\} = \{+,-,+\}$ e.g., [12]. There is another source with $\{-,+,+\}$ [11] [and more [2]. [The occasional $(-\gamma^i)$ convention (e.g., [12]) would reverse the signs on the $\vec{\sigma}\cdot\vec{p}$ terms].

The Dirac equation with Weyl basis still has the form $i{\gamma_c}^\mu\partial_\mu\psi(x) = m\psi(x)$ which in 2x2 matrix form looks like this [2] [12]:

$$H\phi_{Weyl} = \begin{pmatrix} -m & \vec{\sigma}\cdot\vec{p} \\ -\vec{\sigma}\cdot\vec{p} & -m \end{pmatrix}\begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix} = E\begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix}, \tag{25}$$

The bispinor for Dirac and the Weyl or chiral spinors may also be variously labeled as:

$$\psi_D = \begin{pmatrix} u_R \\ u_L \end{pmatrix} = \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}, \; and \; \psi_w = \begin{pmatrix} \chi_L \\ \phi_R \end{pmatrix} = \begin{pmatrix} \varphi_L \\ \varphi_R \end{pmatrix} \tag{26}$$

The Dirac equation for massless particles may also be written in the Weyl basis with Chiral gamma [21] as:

$$\begin{pmatrix} 0 & E - \vec{\sigma}\cdot\vec{p} \\ E + \vec{\sigma}\cdot\vec{p} & 0 \end{pmatrix}\begin{pmatrix} \varphi_L \\ \varphi_R \end{pmatrix} = 0 = \not{p}\psi. \tag{27}$$

The result $\varphi_L = \chi$ and $\varphi_R = \phi$ is the same as above and are considered in a stacked spinor. We also note the shortened "Feynman slash" notation where $\not{p} = \gamma^\mu p_\mu$.

Still in the Weyl basis, the gamma-5 operator gives:

$\gamma_5\psi = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \varphi_L \\ \varphi_R \end{pmatrix} = \begin{pmatrix} -\varphi_L \\ \varphi_R \end{pmatrix}$. And then $P_L = (\frac{1}{2})(1-\gamma_5)\psi = \begin{pmatrix} \varphi_L \\ 0 \end{pmatrix}$.

That is, $\gamma_5\varphi_L = -\varphi_L$ and $\gamma_5\varphi_R = +1\varphi_R$.

In contrast, the Dirac basis γ_5 (an anti-diagonal exchange matrix) operating on the Dirac bispinor acts to exchange the positions of $\psi_R \; and \; \psi_L$, and then $P_L = (\frac{1}{2})(1-\gamma_5)\psi$ mixes them with subtraction.

The standard Dirac equation is appropriate for massive fermions $m > 0$, and the Dirac spinor $\psi = (\psi_R, \psi_L)$ is reducible as the direct sum of two irreducible representations of the Lorentz group. The Dirac spinor is a 4-component spinor that preserves parity while its

[2] With respect to the chiral basis shown in equation (24),Weinberg [5] uses a basis of $\gamma^0 \to -i\gamma^0$, $\gamma^i \to -i\gamma^i$, $\beta = i\gamma^0$, *but* $\gamma^5 = -\gamma^5$

stacked individual Weyl spinors do not. For Lorentz group representations, u_R is in $(\frac{1}{2},\,0)$, u_L has $(0,\,\frac{1}{2})$ and the full Dirac spinor lies in the mixed representation $(\frac{1}{2},\,0) \oplus (0,\,\frac{1}{2})$ [20].

The Dirac 4-spinor can be a stack of 2-spinors of opposite chirality. $\Psi = (\psi_R; \psi_L)$. In the chiral representation, its Lorentz transformation is a block diagonal matrix. We consider only Lorentz boosts of velocity, v, or rapidity, ρ,of the form eqn.(9) and factor out the leading $cosh(\rho/2)$ term to get [12]:

(28)
$$\Lambda_c(v) = \begin{pmatrix} \Lambda(v) & 0 \\ 0 & \Lambda(-v) \end{pmatrix} = cosh(\rho/2) \begin{pmatrix} I - \vec{n}\cdot\vec{\sigma}\, tanh(\rho/2) & 0 \\ 0 & I + \vec{n}\cdot\vec{\sigma}\, tanh(\rho/2) \end{pmatrix}$$

But $cosh(\rho) = \gamma = E/m$ in special relativity. Then we can use: $cosh(\rho) = 2cosh^2(\rho/2) - 1 = 2sinh^2(\rho/2) + 1$ so that $cosh(\rho/2) = \sqrt{(E+m)/2m}$,
$tanh(\rho/2) = \sqrt{(E-m)/(E+m)} = p/(E+m)$. So now in block diagonal form:

$$\Lambda_c = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} I + \frac{\vec{n}\cdot\vec{\sigma}}{E+m} & 0 \\ 0 & I - \frac{\vec{n}\cdot\vec{\sigma}}{E+m} \end{pmatrix} \tag{29}$$

If instead, we had processed the Dirac equation using the standard set of gammas then we would get a Lorentz boost that is not block-diagonal [11]:

$$\Lambda_D(p) = \begin{pmatrix} cosh(\rho/2) & \vec{\sigma}\cdot\vec{n}\, sinh(\rho/2) \\ \vec{\sigma}\cdot\vec{n}\, sinh(\rho/2) & cosh(\rho/2) \end{pmatrix} = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} I & \frac{\vec{\sigma}\cdot\vec{p}}{E+m} \\ \frac{\vec{\sigma}\cdot\vec{p}}{E+m} & I \end{pmatrix}. \tag{30}$$

Applying this momentum $\vec{p}$ boost to a spin up electron in the rest frame, $u(0, s_1)$ yields the "Feynman spinor" $u(p_1, s_1)$ shown below, equation (31).

The chiral representation spinor may be converted into a Dirac representation spinor for a standard view. Remember that a massless case decoupling of the Dirac equation in standard representation required adding and subtracting the coupled equations (eqns.(19)). We called the difference χ_L and the sum ϕ_R. So, the chiral Λ_c transforms the spinor $(\phi_R; \chi_L)$ which can be put back into the $\Psi_D = (u_A; u_B)$ form using $u_A = (\phi_R - \chi_L)\, and\, u_B = (\phi_R + \chi_L)$. It's a little convoluted, but it works. So, the Dirac equation may use either representation.

"Feynman rules" for fermions using spinor formalism enable calculations of scattering cross-sections, decay rates, and radiative corrections. "Parity conserving theories such as QED and QCD are well suited to the four-component fermion methods" [24]. Feynman diagrams with Feynman rules work with a momentum representation (Fourier transform of space-time in $\Psi(x)$).

Assembling the above concepts, suppose we have a simple case of an electron (path 1, spinor labeled $u(p, s)$ for momentum and spin) coming into a scattering vertex and going

out on say path 3 (adjoint spinor $\bar{u}(p,s)$). [There are many different conventions for labeling path legs in Feynman diagrams]. This gets assembled into a vertex vector form: $Vertex_1 = \bar{u}(p_3,s_3) i g_e \gamma^\mu u(p_1,s_1)$ [21] with appropriate coupling $g_e \propto charge\ e$. For a positron, the ingoing particle has an adjoint label $\bar{\nu}(p,s)$ moving backwards in time; and its outgoing state is $\nu(p,s)$.
From the 2×2 form of the Dirac equation in Dirac basis (18), we obtain two coupled equations for massive leptons. As example, the second of these is $\sigma\cdot\vec{p}\,\phi = (E+m)\chi$ so that $\chi = (\sigma\cdot\vec{p}/[E+m])\phi$; and suppose matter-electron $E>0$ so [22]. Then,

$$\text{(31)} \qquad e.g.,\ Feynman\ spinor\ \ u(p_1,s_1) = \begin{pmatrix} \phi_1 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} \phi \\ \frac{\sigma\cdot\vec{p}}{E+m}\phi \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \frac{p_z}{E+m} \\ \frac{p_x+ip_y}{E+m} \end{pmatrix}.$$

$\sigma\cdot\vec{p}$ has the same form as the previous $X = \sigma\cdot\vec{x}$ in eqn. (3). So we are now expressing a Dirac electron wavefunction in terms of spin-up, energy and momentum. And similar examples are easily found for the other lepton paths. For high boosting (p_z becoming relativistic), the spinor acquires a degree of anti-matter-ness.

Two incoming particles "1 and 2" can scatter off of each other due to an inner virtual photon exchange between their scattering vertices (e.g., Møller scattering). That entails a product of the two vertex forms [Vertex 1 times Vertex 2] times a photon"propagator" subject to conservation of momenta expressed via delta functions (a forced constraint). Scattering results in momentum transfer (e.g., $\vec{q} = \vec{p_1} - \vec{p_3}$) from electrostatic 1/r potentials. We are in momentum space, so we must evaluate the propagator as a radial Fourier Transform: $\int \frac{d^3x e^{i\vec{q}\cdot\vec{r}}}{4\pi r} = \frac{1}{q^2}$

This integration can be done just using simple calculus. The resulting scattering cross sections from this process can be seen in many textbooks [10][21][22].

Dirac algebra introduces a new term called the "Dirac adjoint" $\bar{\Psi} = \Psi^\dagger\gamma^o$. This is required to have Lorentz covariant objects that can be formed from a Dirac spinor and its adjoint. $\bar{\psi}\psi$ is a Lorentz scalar and $\bar{\psi}\gamma^\mu\psi$ is a Dirac vector. That means that these joint bilinears remain as scalars or vectors under Lorentz transformations. There are five types of irreducible Lorentz objects: the scalar and vector along with three others – pseudo-scalar, pseudo-vector, and antisymmetric tensor. In the Dirac representation, the Lie algebra $so(1,3)$ acts on $\sigma^{\mu\nu} = i(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)/2$ where the bilinear $\bar{\psi}\sigma^{\mu\nu}\psi$ transforms as a tensor. The term "pseudo-" corresponds to bilinears containing $\gamma_5 = \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ where γ_5 changes sign under a parity transformation (like mirror reflection). Parity on the Dirac spinor is defined by $P:\ \psi(\vec{x},t) \to \gamma^o\psi(-\vec{x},t)$ [20]. It exchanges right-handed and left-handed spinors $u_\pm \to u_\mp$.

7. MATHEMATICAL APPENDIX

Useful concepts from Linear Algebra

An "algebra" consists of a vector space V over a field K together with a law of composition or product of vectors such that scalars a,b,c $\in$ K and vectors A,B,C $\in$ V, A(aB+bC)= aAB + bAC and (aA + bB)C = aAC+bBC.
$A \approx B$: Matrices are "similar" if there is a Hermitian similarity transformation between them: $B = P^{-1}AP$. Then $det(B) = det(A)$ and $tr(B) = tr(A)$. Matrices are diagonalized via similarity transformations.

Hermitian conjugate $A^H = (\bar{A})^T = (A^T)^* = A^\dagger$
A unitary transformation, U, obeys $U^\dagger = U^{-1}$.

From a matrix, A, to its diagonalized matrix, D, a "transformation matrix" (change of basis matrix), P, is formed from columns of eigenvectors ν_i so that $P = [\nu_1, \nu_2, \ldots \nu_n]$. For the diagonalize matrix: $D = P^{-1}AP = diag(\lambda_1, \lambda_2, \ldots \lambda_n)$ where the $\lambda_i{}'s$ are eigenvalues. Also $det(D) = \Pi\lambda_i$ and $A = PDP^{-1}$. Quantum mechanics requires that eigenvectors be normalized so that $\nu^\dagger\nu = I$.
Diagonalizing the 2×2 Pauli matrix σ_x [eqn (2)] results in the diagonal matrix σ_z whose columns are the spinors for electron spin "up" and "down".

For matrices possessing eigenfunctions and eigenvalues: "two eigenvectors of a Hermitian operator corresponding to two different eigenvalues are orthogonal – linearly independent.

Dirac Bar $\bar{\Psi} = \Psi^\dagger\gamma^o$.
$R(\theta)\vec{v}$: A column vector v can be rotated into another vector u using a square rotation matrix. R is a rotation matrix if and only if $R^T = R^{-1}$ and $det(R) = 1$. For 2 or 3-space, $R \in SO(2)\ or\ SO(3)$ — special orthogonal groups.

Hamilton's Quaternions, H, from 1843

Quaternions use three complex numbers {base: 1, i, j, k} and have a long history of use prior to Pauli's electron spin matrices; so they had conventions separate from Pauli and didn't know about electron spin. The 2x2 matrices for quaternions are much less standardized, and there are many representations. One that is common is expressed with respect to Pauli by $e_m = i\sigma_n$ where n can differ from m. Wolfram [1], for example, starts with $e_1 = i\sigma_3$, $e_2 = i\sigma_2$, $e_3 = i\sigma_1$ (not in cyclic permutation order) with the property $e_1e_2 = e_3$ along with Hamilton's desire that $i^2 = j^2 = k^2 = ijk = -1 = e_1e_2e_3$. Some use a convention of $e_m \equiv i\sigma_m$, but this is non-standard and has $e_1e_2 = -e_3$ and $e_1e_2e_3. = +1!$ Instead of the confusing i, j, k, some use h, j, k to differentiate the types of i and/or use i' for the complex imaginary.

The best current convention is that of Misner-Thorne-Wheeler [4] [6] defining the Pauli matrices as $\sigma_k = ie_k$. After all, quaternions came first and Pauli matrices were complexified above them. So $e_k = -i\sigma_k$ is the convention we will use here: shown below. Note that an SU(2) rotation $u(\theta) = exp[-i(\vec{\sigma}\cdot\hat{n}\theta/2]$ really means $e^{+\vec{q}\cdot\hat{n}\theta/2}$ where q is a quaternion.

The matrix representation for quaternions is:

$$e_o = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ e_1 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \ e_2 = \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix}, \ e_3 = \begin{pmatrix} -i & 0 \\ 0 & +i \end{pmatrix}. \tag{32}$$

A quaternion vector $\vec{q}$ is expressed using half-angles $\theta/2$ and becomes negative after a full 2π turn.

Quaternions are the natural basis of the Lie algebra su(2) [e.g., $u_1 = -e_1, u_2 = e_2$, $u_3 = -e_3$]. The Lie algebra $sl(2, C) = su(2) + i\, su(2)$. This complexified Lie algebra is spanned by the H, E, F matrices of eqn.(10) above as: $H = u_3/i, E = (u_1 - iu_2)/2i, F = (u_1 + iu_2)/2i$ which are similar to $J_z, J_+,\ J_-$ operators for angular momentum.

Vectors versus Versors

A "quaternion" is a 4d quantity of the form $q = a_0 + ia_1 + ja_2 + ka_3 \in H =$ a scalar plus an imaginary triplet vector. Hamilton $H = span(1) \oplus span(i, j, k)$. It is a common practice to treat pure quaternions (using just triplet basis i j k without scalar a_o) as vectors. But quaternions have different symmetry properties from vectors. Quaternion multiplications by i j k are CCW rotations of $\pi/2$, *e.g.*, $r = a + ib \longrightarrow ir = ia - b$, and $iir = iia - ib = -a - ib = -r$). These 90^o rotations are called "versors" – or quaternions with $|q| = 1$.

Gibbs/Heavyside vector analysis used the same labels i j k rotational versors of quaternions but re-defined them unit polar vectors. [A polar (or "true") vector reverses sign when the coordinate axes are reversed ($v = a^i e_i \longrightarrow -a^i(-e_i)$) 's] – they stay the same under mirror reflection (up is still up). An axial vector or pseudo-vector flips sign under reflection (e.g., left $\longrightarrow$ right). A pure quaternion is only a vector if i j k are treated as unit vectors. Multiplication of pure quaternions is still a quaternion, but vector multiplication C = A x B is an axial vector (not in the family of true vectors).

Clifford Algebra

a Clifford algebra, $C\ell(V, Q)$, is an algebra generated by a vector space (e.g., $V = R, C$) with a quadratic form (e.g., a metric, Q, or 2nd degree polynomial). It can be extended to include the hypercomplex number systems (such as H). If V has n dimensions, the Clifford algebra can also be symbolized by $C\ell_n(V)$.

A Clifford product ("geometric product" [1878]) is defined as a scalar symmetric product plus an anti-symmetric product (a bi-vector):

$$ab = a \cdot b + a \wedge b = (ab + ba)/2 + (ab - ba)/2. \tag{33}$$

The wedge product ("$\wedge$ "or "exterior product" or "outer product") is a generalization of the usual vector cross product (which itself is only defined in 3d space). From the definition, it follows that $a \wedge b = -b \wedge a$ (and in 3d, $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$). The Clifford product is similar to pure quaternion multiplication $ab = a \cdot b + a \times b$, or angular momentum $rp = r \cdot p + iL$ [all on quaternion ijk vector basis]. Pauli algebra has a similar product:
$(\vec{\sigma} \cdot \vec{A})(\vec{\sigma} \cdot \vec{B}) = \vec{A} \cdot \vec{B} + i\vec{\sigma} \cdot (\vec{A} \times \vec{B})$

One might also begin by expressing the anti-commutator for Clifford generators $\{e_i, e_j\} = 2\delta_{ij} = e_i e_j + e_j e_i$. For $e_1 \perp e_2$, $e_1 \cdot e_2 = 0$, and $e_1 e_2 = -e_2 e_1 = e_1 \wedge e_2 = -e_2 \wedge e_1$. So, for sigma matrices, $\{\sigma_i, \sigma_j\} = 2\delta_{i\,j}\, I_2$, and for gamma matrices $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} I_4$ where eta is the space-time metric and I_4 is a 4x4 unit matrix.

Spinor:

Roger Penrose [1] intuitively defines a spinor as an object which turns into its negative after a complete $2\pi = 360^o$ rotation ; and the action of rotation on a spinor is always double-valued. General spinors were discovered by Elie Cartan in 1913. A spinor is more than just a complex column matrix or vector, and the mathematics of spinors is very difficult. Spinors are the irreducible representations of the 'Clifford group' [11]. The 4-D Dirac Spinor is the bispinor in the plane-wave solution of the free Dirac equation, and a bispinor is the stacking of two Weyl spinors on top of each other in a column matrix. A famous mathematician (Atiyah) said, "No one fully understands spinors. Their algebra is formally understood, but their general significance is mysterious. In some sense they describe the 'square root' of geometry and, just as understanding the square root of -1 took centuries, the same might be true of spinors." A complex 2-D spinor (α, β) represents the fractions of spin up and spin down, $\alpha|\uparrow\rangle + \beta|\downarrow\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$.

The Mathematical Definition of Spinors

Spinor representations are the irreducible representations ("irreps") of the Clifford group obtained from the irreducible representations of the Clifford algebra and its even subalgebra.
"The mapping into the endomorphism algebra of any minimal left ideal induced by the regular representation is called the spinor representation of the simple Clifford algebra and the minimal left ideal is called the space of spinors [13]." "In physics, elements of the vector space carrying an irreducible representation of the complexified Clifford algebra are termed "Dirac Spinors." For rotations: Spinors are objects which carry an irreducible representation of the spin group which is the double cover of SO(3) (e.g) and is the spin 1/2

representation of the group of rotations in a quadratic space.

Names for the types of spinors include: [25]
"Classical spinors" based on irreps of $Spin_+(p,q)$ [*e.g.*, $Spin_+(1,3) \simeq SL(2,C)$. Note that this "severely restricts the analysis to the usual Dirac, Weyl, and Majorana spinors."]
These are sections of the vector bundle $P_{Spin_{1,3}} \times C^2$.
"Algebraic spinors" have Clifford algebra irreps (minimal left ideal). And "Operational spinors" have Clifford algebra using the representation space associated to the even subalgebra.

Dotted versus UnDotted spinor notation:

We need to distinguish components with Weyl "left" handed representation from those with right or Dirac mixed representations: $\ell = (\frac{1}{2}, 0)$ while the "right" Weyl spinor is in the Lorentz representation $r = (0, \frac{1}{2})$, and general 4d Dirac fermion representation is $(\frac{1}{2}, 0) \oplus 0(\frac{1}{2}) = \ell \oplus r$ – also called a bispinor representation. This separation is done using "Van der Waerden dotted" notation where Weyl ℓ spinors have undotted chiral indices (like $\alpha,\ \beta, \ldots$ where $\alpha = 1, 2, \ldots$) and r spinors have dotted indices $\dot{\alpha},\ \dot{\beta}$ where $\dot{\alpha} = \dot{1}, \dot{2}, \ldots$.

In the "ℓ" representation, we have Lorentz transformation group elements $g = e^{M_\ell}$ where the exponent matrix $M_\ell = [(\rho/2 - i\hat{n}\theta/2) \cdot \vec{\sigma}]$. If we have a two-component ℓ-spinor ψ_α, it transforms as $\psi_\alpha \longrightarrow {M_\alpha}^\beta \psi_\beta$ [24].
For the r-representation, we use the complex conjugate matrix M*. ℓ and r are Hermitian conjugates, so the transformation would be ${\psi^\dagger}_{\dot{\alpha}} \longrightarrow {M^*}_{\dot{\alpha}}{}^{\dot{\beta}} {\psi^\dagger}_{\dot{\beta}}$.
"For typographical reasons, Penrose replaced the dotted indices with primed indices, a notation still employed by most general relativists today" [24].

For physicists, another convention is given by the favored text Gravitation [4] under the chapter on spinors.
Spinors with raised dotted indices plus an overbar on the symbol are RH and called anti-chiral. Indices with hats are Dirac indices like $A = 1, 2$ or dotted $\dot{A} = \dot{1}, \dot{2}$.

A Lorentz transformation of a spinor $\xi' = L\xi$ is more carefully labeled (using capital letters) as ${\xi'}^A = {L^A}_B \xi^B$, and this uses half of the transformation formula $Q' = LQL^*$. Then one introduces another spinor η transforming by "the conjugate complex of the Lorentz transformation" : ${\eta'}^{\dot{U}} = \bar{L}^{\dot{U}}_{\dot{V}} \eta^{\dot{V}}$.

Irreducible and Spinor Representations

This is an important, and very difficult topic. An outline of it is the first three chapters of the book, Group Theory and General Relativity, by Moshe Carmeli [15]. It is too lengthy to present here.

Irreducible representations for spaces of quantum angular momentum is discussed in the text by Messiah [18]. A "standard representation" is called $\{J^2 J_z\}$ in which the z'th component of angular momentum is diagonal and J^2 is also diagonalized. If the ket-state $|jm\rangle$ is an eigenvector of these, then $J^2|jm\rangle = j(j+1)|jm\rangle$ and $J_z|jm\rangle = m|jm\rangle$. For a selected value of total angular momentum, j, there are $2j+1$ values of m, the z'th projection, with integer spacings: $m = -j, -j+1, -j+2, \ldots, +j-1, +j$. A raising operator can be found taking the projection m up to its next highest value $m+1$.

$$J_+|jm\rangle = \sqrt{j(j+1) - m(m+1)}|jm\rangle = \sqrt{(j-m)(j+m+1)}|jm\rangle \tag{34}$$

Starting with $m = -j$ there can be 2j applications of a "raising operation" or "ladder" increasing m up to $+j$. The series of $2j+1$ $|jm\rangle$ ket-vectors correspond to a subspace $E^{(j)}$ of a total Hilbert space. This subspace is invariant under rotation or "irreducible" with respect to rotations.
Finding irreducible representations of groups is important "because the basic fields of physics transform as irreducible representations of the Lorentz and Poincare groups [11]."
Rank of a Spinor:

In physics, rank n usually means how many Weyl chiral spinors in a tensor product. "Equivalently, it is the total number of dotted and undotted indices on the spinor." For a $3+1$ irreducible representation with left and right spins (j_L, j_R), $n = 2j_L + 2j_R$. So a single Weyl spinor is called rank 1. Spinors of rank 1 may also be labeled as rank (1,0) or (0,1) for one undotted or one dotted index.

In the previous discussion, a spinor is rank one transforming as $\vec{s'} = \Lambda\vec{s}$ under a change in inertial frame. The outer product $ss^\dagger$ is a 2x2 matrix 2nd rank spinor transforming as $ss' \longrightarrow \Lambda ss^\dagger\Lambda^\dagger$. The previous matrix Q (eqn.(8)) is a 2nd rank spinor, $Q' = \Lambda Q\Lambda^\dagger$.

Algebra, Rings, Groups:

A ring $\langle A, +, \cdot\rangle$ is a set and binary operations that are an abelian group under "$+''$, associative for $\cdot$, $\cdot$ is left and right distributive with respect to "$+''$.
A subgroup $K \subset A$ is a "left ideal" when: $K \neq \emptyset$, $x, y \in K \Longrightarrow x - y \in K$, $\forall x \in K, a \in A, ax \in K$ [26].
A homomorphism of $\langle A, \circ\rangle$ into itself is an endomorphism. An isomorphism of $\langle A, \circ\rangle$ into itself is an automorphism.
A map $A \mapsto A'$ is an isomorphism if it is bijective.
For a subgroup $H \subset G$, left coset is an equivalence class $a \sim b$ of an element $a \in G$ if $b = ah$ for some $h \in H$ [13].
(E.g., see "The mathematical definition of spinors" in the section on spinors).
Symplectic matrix is real $2n\times 2n$ such that $M^T\Omega M = \Omega$ (follows that $M^{-1} = \Omega^{-1}M^T\Omega$).
A frequent choice is $\Omega = [0,\ I_n;\ -I_n,\ 0]$ or upper-triagonal, lower-triagonal, and I with entries 0 and 1. $det(\Omega) = +1$, $\Omega^{-1} = \Omega^T = -\Omega$.

The Center of an algebra A is $Z(A) = \{a \in A | ab = ba \forall b \in A\}$.

Lie Groups and Lie Algebras:

The Lie algebra of a Lie group is the first linear approximation of the group (the tangent space about the group identity, I).It is a vector space $\mathfrak{g}$ over a field (like R or C) along with a bilinear map that keeps the outcome in the vector space: $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$. The map is called the commutator or "Lie bracket" $[a, b] = ab - ba$, $[a, a] = 0$ (anti-commutativity). These brackets are incorporated into a requirement called the Jacobi identity: $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$. Roughly, this is an analog of associativity for infinitesimal symmetries. The dimension of a Lie algebra is its dimension as a vector space over the field.
Lie (continuous) Groups use capital letters, and their Lie Algebras us small letters (ideally "Fraktur" letter font like $\mathfrak{g}$). For a Lie Group, G, the Lie algebra g is the tangent space of G at the identity, I (infinitesimals around 1). There is a surjective exponentiation map $exp: \ g \to G, \ M \mapsto exp(M)$

"The general linear group of degree n is the set of n x n invertible matrices, together with the operation of ordinary matrix multiplication. This forms a group, because the product of two invertible matrices is again invertible, and the inverse of an invertible matrix is invertible. The group is so named because the columns of an invertible matrix are linearly independent" [8]. GL(2,C) is the group of linear transformations on C^2 which are invertible. Another way of looking at it is all complex 2x2 matrices with non-zero determinant. For SL(2,C), the determinant is one.

One definition for a Lie algebra is:
$Lie(G) = \{m \in M(n, C) \,|\, \forall t \in R, exp(tm) \in G \subset GL(n, C)\}$.

Relevant Groups:
$SU(2) = \{M \in GL(2, C) | \, MM^\dagger = 1 = M^\dagger M, \ det(M) = 1\}$ [i.e., Unitary].
$su(2) = \{M \in gl(2, C) | M + M^* = 0, \ tr(M) = 0\}$.
$SL(2, C) = \{M \in GL(2, C) | \, det(M) = 1\}$.
$sl(2, C) = \{M \in gl(2, C) | \, tr(M) = 0\} = su(2) \oplus \ i \, su(2)$.
$sl(2, C) = span\{E, \ F, \ H\} = z_1 E + z_2 F + z_3 H$,
$su(2) = span\{i \, , j, \ k\}$, [or in matrix form eqn.(4), set $a = (I + a_{Im})$.
$H = span(I) \oplus span\{i \, , j, \ k\}$.

A "spin group" is a double cover of the corresponding $SO(p, q)$.
So, $spin(2) = U(1) \cong SO(2)$.
$spin(3) \cong SU(2) \cong SO(3)$, $spin(1, 3) \cong SL(2, C) \cong SO^+(1, 3)$

The group SU(2) has the same Lie algebra as SO(3), i.e., $[X_i, X_j] = i\epsilon_{ijk} X_k$,
e.g., $\sigma_1 \sigma_2 - \sigma_2 \sigma_1 = 2i\sigma_3$ with $X_i = \sigma_i \, /2$.

Spin And Statistics:

"Quantum mechanics says that if you turn a particle around 360^o, its wavefunction changes by a phase of either +1 (that is, not at all) or -1. It also says that if you interchange two particles of the same type, their joint wavefunction changes by a phase of +1 or -1.

The Spin-Statistics Theorem says that these are not independent choices: you get the same phase in both cases! The phase you get by rotating a particle is related to its spin, while the phase you get by switching two" is called statistics [29]. This means that
i. Physical systems that obey Bose/Einstein statistics possess integer spin; and
ii. Physical systems that obey Fermi/Dirac statistics possess half-integer spin.

In non-relativistic quantum mechanics, a two-particle wavefunction has the addition of the original and exchanged wave function,
e.g., $\Psi = (1/\sqrt{2})[\psi(1,2) + \psi(2,1)]$, a permutation. The exchanged $\psi(2,1) = (-1)^{2s}\psi(1,2)$ saying that a fermion spin $s = 1/2$ exchanged function acquires a minus sign in its "spin factor" giving an anti-symmetric Ψ; but for bosons the factor is +1.

Another way to say this is the vanishing of commutator or anti-commutator relationships (subtracting versus adding):
$[\psi_1(x), \psi_2(x')] = 0$ or $\{\psi_1(x), \psi_2(x')\} = 0$. And $\{\psi_1(x), \psi_2(x')\} = 0 \Longrightarrow [\psi_1(x), \psi_2(x')] \neq 0$ applying to vector or spinor fields. And fermion particles with half-integer spin identify with the spinor representation of the rotation group.

The term "exchange" is defined mathematically as just a permutation – but there must be some physics taking place also.

Perhaps the best example is the scattering of two particles in the center of mass CM system discussed in the Feynman lectures[32]. A particle comes in from the left and another from the right meeting at the center. The left particle interacts and scatters through an angle θ meaning that the other particle must scatter at the angle $\pi - \theta$. There are two detectors: an upper one D_1 and an opposite one below D_2 – we only need to consider the probability P of some particle detected in D_1.

For distinguishable particles (alpha on oxygen nucleus or for two electrons with opposite spins), P is just the sum of squares: $P = |f(\theta)|^2 + |f(\pi-\theta)|^2$. For boson identical particles, we can't tell which particle went where and $P = |f(\theta) + f(\pi - \theta)|^2$. For identical fermions, we subtract amplitudes $P = |f(\theta) - f(\pi - \theta)|^2$.
For the special state $\theta = \pi/2 = 90^o$, $f(\theta) = f(\pi - \theta) = f(\pi/2) = f$. The probability P for the different cases is then very interesting: P(distinguishable) $= 2f^2$ versus P(identical bosons) $= 4f^2$ versus P(identical fermions) $= 0$! The fermion case gets nullified! And, "it is twice as likely to find two identical Bose particles scattered into the same state as you

would calculate assuming the particles were different." This enhancement ultimately led to the laser.
Expressed as two particle states, left and right to detectors 1 and 2, $P_2 = |\ell_1 r_2 - r_1 \ell_2|^2$.

Quantum mechanics is a theory of measurements with respect to preparation and detection at a distance. And exchange expresses possibilities viewed from a distance. The interaction region doesn't know which two-particle state gets the minus sign, and it doesn't matter because use of the Born rule "squares" amplitudes to get probabilities.

In Quantum Field Theory, QFT, particle number need not be conserved; and we switch focus to particle creation and annihilation operators. We say that the symmetric state for bosons and the anti-symmetric state for fermions with momentum p obey:
$[a(\vec{p}),\, a^\dagger(\vec{p}\,')] = \delta(\vec{p} - \vec{p}\,')$, and $\{a(\vec{p}),\, a^\dagger(\vec{p}\,')\} = \delta(\vec{p} - \vec{p}\,')$.

Now, in quantizing the Dirac field, we have to address both matter and antimatter together (a 4-spinor). Essentially, the Dirac Hamiltonian has terms that look like the first part of the equation below with b operators for matter and d operators for antimatter (the subscript s is the spin state) [11]. The problem with these terms in the equation is that the second term with the minus sign implies that energy is unbounded from below (the system energy could be negative! – deemed "not sensible"). This is solved in two steps. The first is called "Normal ordering" which says that particles must always be created (dagger) before being destroyed (the dagger term must come first). "The vacuum expectation value of a normal ordered product of creation and annihilation operators is zero" where $|0\rangle$ represents the vacuum, and operator $a|0\rangle = |0\rangle$. This avoids zero-point energy. The second is that anticommutation rules must be enforced for fermions and for antifermions so that $\{d_s, d^\dagger{}_s\} = 0$. We then get:

$$(b_s{}^\dagger(p) b_s(p) - d_s(p) d_s{}^\dagger) \longrightarrow (b_s{}^\dagger(p) b_s(p) + d_s{}^\dagger(p) d_s(p)) \tag{35}$$

So, energy positivity is one motivation for requiring fermi statistics and its subsequent exclusion principle. Spin-Statistics also requires relativistic Lorentz invariance, three spatial dimensions, and relativistic causality ["microcausality is the requirement that two physical measurements made in different points x and y be mutually independent, if these two measurements were made with spatial distance"]. "Quantum field theory enforces the connection between spin and statistics" [22].

The spin-statistics theorem for non-relativistic quantum mechanics is empirical and postulated. The resulting "Pauli Exclusion Principle" (PEP) for antisymmetric wave functions is extremely important and helps give matter its rigidity (holds up mountains, white dwarfs, and neutron stars). But formal proofs of the SS theorem require quantum field theory [Relativistic QFT] and are deemed to be difficult, unclear and not quite valid. There are hundreds of papers on proofs that fall into two general approaches [31]: A formal rigorous purist "Wightman or Algebraic" approach has a problem of not producing "realistic interacting models of the relevant axioms." The pragmatist "Lagrangian or Weinberg" approach is intuitively clearer but not rigorous: it uses power series expansions of the S-matrix, contains divergent terms at high energies and has problems with convergence. The Lagrangian approach was first due to Fierz and Pauli in 1940 [30]. But it is "backhanded": bosons cannot have Fermi-Dirac statics, and fermions cannot have Bose-Einstein statistics. Straightforward clarity remains elusive. The encompassing problem is that "There is no unique set of first principles from which SS can be derived in RQFTs" [31]. Although the Pauli exclusion principle dates back to 1925, its lack of clear or firm understanding should place it on the frontier of physics as an important and outstanding problem still needing to be solved.

A comprehensive reference agrees [33]: "What is proved . . . is that the existing theory is consistent with the spin-statistics relation. What is not demonstrated is a reason for the spin-statistics relation" . . . "The spin-statistics theorem could conceivably be an essential ingredient of a more fundamental view of the world."

References

[1] Roger Penrose, The Road to Reality, Alfred A. Knopf, N.Y., 2005
[2] William O. Straub, "Weyl Spinors and Dirac's Electron Equation," 2005, www.weylmann.com/weyldirac.pdf, or http://www.weylmann.com/why.shtml
[3] William O. Straub, "A Child's Guide to Spinors," 2016 viXra:1611.0212 and at Welymann.com.
[4] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler, GRAVITATION, W. H. Freeman and Company, San Francisco, 1970−1973.
[5] Steven Weinberg, The Quantum Theory of Fields, Vol 1 Foundations, Cambridge, 1995
[6] Dave Peterson, "Properties of Biquaternions," 1976
[7] Dave Peterson, "Electron Spin and SU(2)," Oct.2012.
[8] Quaternions in Wikipedia and also Wolfram convention. Other pages on Euler angles. etc.
[9] Theodore Frankel, The Geometry of Physics, An Introduction, Second Edition, Cambridge, 2004.
[10] Frances Halzen, Alan D. Martin, Quarks and Leptons, An Introductory Course in Modern Particle Physics, Wiley, 1984.
[11] Michio Kaku, Quantum Field Theory, Oxford Press, 1993.
[12] Andrew M. Steane, "An Introduction to Spinors," arXiv:1312.3824k [math-ph] 13 Dec 2013. 23 pages.
[13] I. M. Benn, Robert W. Tucker, An Introduction to Spinors and Geometry with Applications in Physics, Adam Hilger, Bristol and New York, 1987 (p54, p92). [Mathematically advanced]
[14] G. Costa and G. Fogli, Symmetries and Group Theory in Particle Physics, Lecture Notes in Physics, Springer-Verlag Berlin Heidelberg 2012. Chap. 2 "The rotation group."
[15] Moshe Carmeli, Group Theory and General Relativity, McGraw-Hill, 1977.
[16] Dave Peterson, "Lie Group Representations," June 7, 2015.
[17] Seymour Lipschutz, Marc Lipson, Schaum Outline: Linear Algebra, 5th Ed., 2013
[18] Albert Messiah, Quantum Mechanics, Vol. 11, John Wiley and sons, 4th printing 1965.
[19] Jacob K. Goldhaber, Gertrude Ehrlich, Algebra, Macmillan, 1970.
[20] Chris J. L. Doran, "Geometric Algebra and its Application to Mathematical Physics," Cambridge PhD dissertation, 188 pages, Feb. 1994.
[21] David McMahon, Quantum Field Theory DeMystified , McGraw-Hill, 2009
[22] I.J.R. Aitchison, A.J.G. Hey, Gauge Theories in Particle Physics, 2nd Ed., Adam Hilger, 1989.
[23] A. Zee, Quantum Field Theory in a nutshell, Princeton, 2003.
[24] Herbi K. Dreiner, Howard E. Haber, Stephen P. Martin, "Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry," 300 pages. ArxiV hep-ph 0812.1594 .

[25] J.M.Hoff da Silva, Roldao da Rochab, "Unfolding physics from the algebraic classification of spinor fields," Physics Letters B, Volume 718, Issues 4?5, 29 January 2013, Pages 1519-1523 (open access) or https://www.sciencedirect.com/science/article/pii/S0370269312012798 .
[26] H. Goldstein. Classical mechanics. 1951.
[27] University of Gothenberg department of physics, "5 The Dirac Equation and Spinors," physics.gu.se /∼tfkhj/ TOPO/DiracEquation.pdf
[28] Hans de Vries, Understanding Relativistic Quantum Field Theory, Chapter 16 "The relativistic Dirac equation," 2008, http://chip-architect.com/physics/Book_Chapter_Dirac.pdf
[29] John Baez, "Spin, Statistics, CPT and All That Jazz," 2012 http://math.ucr.edu/ home/baez/ spin_stat.html
[30] Wolfgang Pauli, "The connection between spin and statistics." Phys. Rev. Vol. 58, 716, 1940.
[31] Jonathan Bain, "What Explains the Spin?Statistics Connection?" http://faculty.poly.edu/ jbain/ Talks/ExpSSC.pdf 36 slides
[32] Feynman, Leighton, Sands, The Feynman Lectures on Physics, Volume III (Quantum Mechanics), Addison Wesley, 1965.
[33] E.C.G. Sudarshan and I. Duck, Pauli and the Spin-Statistics Theorem (World Scientific, Singapore, 1997)/

PROPERTIES OF BIQUATERNIONS

A biquaternion is a quaternion having complex components: $Z = z_0 + z_1\sigma_1 + z_2\sigma_2 + z_3\sigma_3 = x_0 + \vec{x} + iy_0 + i\vec{y}$ where $z_\mu = x_\mu + iy_\mu$. A real biquaternion is given by $X = x_0 + \vec{x}$. The σ_k's are isomorphic to the Pauli matrices and have the properties that $\sigma_k^2 = 1$ and that $\sigma_i\sigma_j = -\sigma_j\sigma_i = i\sigma_k$ (i,j,k are cyclic permutations of 1,2,3). We define the quaternions by $q_k = -i\sigma_k$ so that $q_k^2 = -1$ and $q_1q_2q_3 = -1$ (or equivalently, $q_1q_2 = q_3$, some authors use $q_1q_2 = -q_3$). Thus, in terms of matrix representations we have:

$$\left[\begin{array}{llll} \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} & \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & \text{often } e_k = \sigma_k \text{ and } e_0 = 1 \\ q_1 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} & q_2 = \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix} & q_3 = \begin{pmatrix} -i & 0 \\ 0 & +i \end{pmatrix} & \text{are used as bases.} \end{array}\right]$$

In ordinary physics, real numbers have priority and complex numbers are used less frequently. Similarly, the real biquaternions have the better physical utility and the quaternions can be added when desirable. It is our contention that Hamilton's greatest error was his preference for his quaternions (the hypercomplex analog of $i = \sqrt{-1}$) over the real biquaternions (which he also discovered).

Def:

$$\left[\begin{array}{ll} \text{Hyperconjugate} & Z^\dagger = x_0 - \vec{x} + iy_0 - i\vec{y} = z_0 - \vec{z}\,. \\ \text{Complex Conjugate} & Z^* = x_0 + \vec{x} - iy_0 - i\vec{y}\,. \\ \text{Quaternion conjugate} & \bar{Z} = x_0 - \vec{x} - iy_0 + iy = (Z^*)^\dagger = (Z^\dagger)^*. \\ \text{Biquaternion Norm} & N(Z) = Z^\dagger Z = Z\,Z^\dagger = z_0^2 - z_1^2 - z_2^2 - z_3^2. \end{array}\right]$$

A real biquaternion has $X^* = X$. Regular quaternions have $\bar{Q} = Q$. Therefore, the operation of quaternion conjugate means "negate that part of a biquaternion which is not a regular quaternion." Regular quaternions obey $q_i q_j = q_k = -q_j q_i$ and $q_k^2 = -1$, so we would expect their multiplication to be different from that of real biquaternions. In particular:

$$\left[\begin{array}{l} QQ' = x_0x_0' + x_0\vec{y}' + x_0'\vec{y} - \vec{y}\cdot\vec{y}' + \vec{y}\times\vec{y}', \text{ and} \\ Q^\dagger Q = |Q|^2 = N(Q) = x_0^2 + y_1^2 + y_2^2 + y_3^2 \quad \text{with } |QQ'| = |Q|\,|Q'|\,. \\[1ex] XX' = x_0x_0' + x_0\vec{x}' + x_0'\vec{x} + \vec{x}\cdot\vec{x}' + i\vec{x}\times\vec{x}', \text{ and} \\ X^\dagger X = |X|^2 = N(X) = x_0^2 - x_1^2 - x_2^2 - x_3^2 \quad \text{with } |XX'| = |X|\,|X'|\,. \end{array}\right]$$

If $N(X) > 0$ then X is said to be timelike and $|X|$ is real. Hyperconjugation obeys the rules $(Z + Z')^\dagger = Z^\dagger + Z'^\dagger$ and $(ZZ')^\dagger = Z'^\dagger Z^\dagger$. The trace of Z is defined to be $Tr(Z) = Z + Z^\dagger = 2z_0$. We can also define an inner product and a metric tensor for biquaternion space by letting:

$$\left[\begin{array}{l} (\alpha|\beta) \equiv \tfrac{1}{2}[\alpha^\dagger\beta + \beta^\dagger\alpha] = g_{\mu\nu}\alpha^\mu\beta^\nu = \alpha^0\beta^0 - \alpha^1\beta^1 - \alpha^2\beta^2 - \alpha^3\beta^3 \\ g_{\mu\nu} = \tfrac{1}{2}(e_\mu^\dagger e_\nu + e_\mu e_\nu^\dagger + e_\nu^\dagger e_\mu + e_\nu e_\mu^\dagger) = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \text{ which is timelike.} \end{array}\right]$$

The norm is then $N(\alpha) = (\alpha|\alpha)$. Physical four-vectors like $x^\mu = (ct, \vec{x})$ and $p^\mu = (E/c, \vec{p})$ can be represented by real biquaternions and have invariant inner products: $(X|X) = c^2\tau^2$, $(P|P) = m_0^2c^2$. Sometimes a basis of i, q_1, q_2, q_3 is used and gives a spacelike inner product so that $(X|X) = -c^2\tau^2$. This basis is called the Minkovskian quaternions or miniquats.

Biquaternions are a non-commutative ring with unity. Unfortunately the division axiom is not satisfied; ie., $ZZ' = 0$ does not require Z or Z' to be zero. Also, multiplicative inverses may not exist. For example, when $N(Z) \neq 0$, inverses are defined by $Z^{-1} = Z^{\dagger}/N(Z)$; but, if $x_0 = ct$, then $N(X)$ will be zero on the light cone. Since $N(Q)$ is always >0, quaternions will have inverses, $Q^{-1} = Q^{\dagger}/N(Q)$. Q is therefore a division ring and is not a field because it is non-commutative. This means that polynomials such as $\alpha^2 + 1 = 0$ could have an infinite number of roots.

Biquaternions are a vector space and an algebra. A linear transformation of a vector space into itself which has nullity zero (dim Ker L = 0) is an automorphism. Automorphisms are the only linear transformations which have inverses and are therefore non singular. We are going to consider elements of Minkowski space to be represented by X. A Lorentz transformation is a linear transformation mapping X onto X which preserves the value of the inner product. The Lorentz group, $L = O(1,3)$, is the group of orthogonal automorphisms of Minkowski space, $\mathbb{R}^{1,3}$. Every automorphism of the quaternions is required by theorem to be of the form $Q' = AQA^{-1}$ where $|A| = 1$ and hence $A^{-1} = A^{\dagger}$. Suppose, for example, that $A = q_1$. Then $A^{\dagger} = -q_1$ and $A(x_0 + y_1q_1 + y_2q_2 + y_3q_3)A^{\dagger} = (x_0 + y_1q_1 - y_2q_2 - y_3q_3)$. Thus $A = q_1$ produces an inversion of the q_2 and q_3 axes. We need to multiply twice on Q because it takes a product of two quaternions to get back to the real numbers. Since $q_K^{\,2} = -1$ and $\sigma_K^{\,2} = +1$, it turns out that the best formula to transform X is $\underline{X' = BXB^*}$,--conjugate instead of hyperconjugate. Thus if $B = \sigma_1$, $B^* = \sigma_1$ and we see that $\sigma_1(x_0 + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3)\sigma_1 = (x_0 + x_1\sigma_1 - x_2\sigma_2 - x_3\sigma_3)$, an inversion of the σ_2 and σ_3 axes. Notice that $B = -i\sigma_1$, $B^* = i\sigma_1$, would accomplish the same thing, a rotation by π radians. The following formula gives a rotation by an arbitrary angle θ about the σ_K axis:

$$\left[\; R_K(\theta) = \cos\tfrac{1}{2}\theta - i\sigma_K\sin\tfrac{1}{2}\theta = \exp(-i\sigma_K\tfrac{1}{2}\theta). \;\right]$$

If we performed an unusual rotation through an imaginary angle $-i\alpha$, we would get:

$$\left[\; L(\alpha) = \cosh\tfrac{1}{2}\alpha - \sigma_K\sinh\tfrac{1}{2}\alpha = \exp(-\sigma_K\tfrac{1}{2}\alpha). \;\right]$$

This turns out to be a pure boost, a Lorentz transformation. For an arbitrary orientation, let $\hat{n}$ (or $\hat{\beta}$) $= \cos\alpha\sigma_1 + \cos\beta\sigma_2 + \cos\gamma\sigma_3$ be a unit vector along the axis of rotation (or along the axis of motion). The general Lorentz transformation is then:

$$\left[\; L(\alpha,\theta) = \exp[-\tfrac{1}{2}(\vec{\alpha} + i\theta\hat{n})], \text{ where } \vec{\alpha} = \hat{\beta}\tanh^{-1}\beta, \beta = v/c. \;\right]$$

If v is in the x direction and we calculate $L_1(\alpha)XL_1(\alpha)^*$ we get. $(ct)' = \gamma(ct - xv/c)$ and $(x)' = \gamma(x - vt)$ as we should. *For a rotation of X by plus θ about the σ_3 axis we obtain:*

$$\left[\; \underline{R_3(\theta)XR_3(\theta)^* = \chi_0 + \sigma_1(\chi_1\cos\theta - \chi_2\sin\theta) + \sigma_2(\chi_2\cos\theta + x_1\sin\theta) + \sigma_3\chi_3} - \;\right]$$

The operation $R_3(\theta)^*XR_3(\theta)$ corresponds to the vector X rotating through <u>minus θ</u>. In matrix form, if we calculate $\chi_\alpha\sigma_\alpha = \Lambda_{\alpha\beta}\chi_\beta\sigma_\alpha$ and compare with $L_3(\alpha)XL_3(\alpha)^*$ we obtain:

$$\chi_0' = \Lambda_{00}\chi_0 + \Lambda_{01}\chi_1 + \Lambda_{02}\chi_2 + \Lambda_{03}\chi_3 = \gamma(\chi_0 - \chi_3\beta), \text{ etc.}$$

<u>Therefore, the operation L_3R_3 corresponds to:</u>

($\cosh\alpha = \gamma$, $\sinh\alpha = \beta\gamma$)

$$\underline{\Lambda_{\alpha\beta} + R_{\alpha\beta}} = \begin{bmatrix} \cosh\alpha & 0 & 0 & -\sinh\alpha \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ \sinh\alpha & 0 & 0 & \cosh\alpha \end{bmatrix}$$

Def: 1. $\vec{\sigma} = \sigma_1 i + \sigma_2 j + \sigma_3 k$.

2. $\vec{\sigma}\cdot\vec{A} = A_1\sigma_1 + A_2\sigma_2 + A_3\sigma_3 = \underline{A}$ = a real vector biquaternion.

3. Cross Product $\underline{A} \times \underline{B}$ = vector part of $\underline{AB} = i(\vec{A}\times\vec{B})\cdot\vec{\sigma}$. The vector $(\vec{A}\times\vec{B})$ is understood to be in a Pauli matrix base.

4. Dot Product $\underline{A}\cdot\underline{B}$ = scalar part of $\underline{AB} = \vec{A}\cdot\vec{B}$.

5. The following formula is used in the quantum mechanics of angular momentum. We now recognize it as simple quaternion multiplication. $(\vec{\sigma}\cdot A)(\vec{\sigma}\cdot B) = A\cdot B + i\vec{\sigma}\cdot(\vec{A}\times\vec{B})$ or

$$\underline{AB} = \vec{A}\cdot\vec{B} + i\vec{A}\times\vec{B} = \underline{A}\cdot\underline{B} + \underline{A}\times\underline{B}.$$

for example, $\underline{rp} = \vec{r}\cdot\vec{p} + i\vec{r}\times\vec{p} = \underline{r}\cdot\underline{p} + i\underline{L}$

6. An axial vector (or pseudovector) reverses sign upon inversion. Any cross product of two vectors is a pseudovector, ie., $\vec{A}\times\vec{B} = -\vec{B}\times\vec{A}$. $\vec{L} = \vec{r}\times\vec{p}$ and $\vec{B} = \nabla\times\vec{A}$ are pseudovectors. The following operation is called the angular momentum axial 4-vector:

$$x\boxtimes p = \tfrac{1}{2}(x^{\dagger}p - p^{\dagger}x) = ct\underline{p} - \underline{x}E/c - i\underline{x}\times\underline{p}.$$

7. A pseudoscalar is a scalar which changes sign under an improper rotation (inversion). The product of a vector and an axial vector is a pseudoscalar. eg., the triple product $A\cdot B\times C$.

8. $\partial^{\dagger} \equiv \left(\frac{1}{c}\frac{\partial}{\partial t}, \nabla\right)$, $\partial \equiv \left(\frac{1}{c}\frac{\partial}{\partial t}, -\nabla\right)$. Using $AB = a_ob_o + a_o\underline{b} + \underline{a}b_o + \underline{a}\cdot\underline{b} + i\vec{a}\times\vec{b}$, we have $\partial^{\dagger}\partial = \partial_o{}^2 - \nabla^2 \equiv \Box$ = the D'Alembertian. We reserve the symbol $\Box^2$ for the usual spacelike operator $\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$.

9. Minkovskian quaternions (min^quats): $M = m_o i + m_1 q_1 + m_2 q_2 + m_3 q_3$. $i = \sqrt{-1}$. Their multiplication obeys :

$MN = -m_o n_o + i m_o \underline{n} + i n_o \underline{m} - \underline{m}\cdot\underline{n} + \vec{m}\times\vec{n}$, and $(m/n) = -m_o n_o + \underline{m}\cdot\underline{n}$, spacelike. Since this system also has a pseudo-Euclidean metric and invariant inner products, it is as adequate as the real biquaternions.

10. A vector $\underline{Q} = a_1q_1 + a_2q_2 + a_3q_3$ with real coefficients will be called a q-vector. A vector $\underline{X} = b_1\sigma_1 + b_2\sigma_2 + b_3\sigma_3$ with real coefficients will be called a p-vector (p for Pauli). A regular quaternion will sometimes be called a quat; and a real biquaternion will sometimes be called a squat (s for sigma).

Comments: $AB^{\dagger} + BA^{\dagger} = B^{\dagger}A + A^{\dagger}B = 2(A/B) = 2(B/A)$ is a Lorentz invariant. ie., $a_ob_o - \underline{a}\cdot\underline{b}$ is a scalar and $L^{-1}z_o\sigma_oL = z_o\sigma_o$. Scalars and space-like numbers are special cases of axial 4-vectors, eg. $\underline{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. These objects do not satisfy the usual Lorentz transformations for proper 4-vectors. Instead they must transform as $a \rightarrow a' = L^{\dagger}aL = a_o\sigma_o + L^{\dagger}\underline{a}L$.

ELECTROMAGNETISM:

Let the electromagnetic 4-potential be $A = (\phi/c, \underline{A})$ and let J = the 4-current, $(c\rho, \underline{J})$. The derivative of potential is:

$$\partial^{\dagger}A = \frac{1}{c^2}\frac{\partial\phi}{\partial t} + \nabla\cdot A + \frac{1}{c}\frac{\partial A}{\partial t} + \frac{1}{c}\nabla\phi + i\nabla\times A.$$

Using the Lorentz gauge we have $\nabla\cdot A + \mu_0\epsilon_0\partial\phi/\partial t = 0$ where $\mu_0\epsilon_0 = 1/c^2$. Also $\underline{E} = -\nabla\phi - \partial A/\partial t$ and $\underline{B} = \nabla\times A$ (MKS units). So: $\partial^{\dagger}A = -\underline{E}/c + i\underline{B} = -(\underline{E}/c - i\underline{B}) = \psi$ a p-vector + a q-vector. We can justify B being a q-vector by noting that $e\underline{v}\times(-i\underline{B}) = -iei(\vec{v}\times\vec{B}) = e\vec{v}\times\vec{B}$, which is real as it should be. If we take the D'Alembertian of A we get:

$$\Box A = \partial^{\dagger}\partial A = \frac{1}{c}\frac{\partial}{\partial t}(-\underline{E}/c + i\underline{B}) - \nabla\cdot(-\underline{E}/c + i\underline{B}) - i\nabla\times(-\underline{E}/c + i\underline{B}).$$

$$= -\frac{i}{c}\nabla\cdot\underline{B} + \nabla\cdot\underline{E}/c + \nabla\times\underline{B} - \frac{1}{c^2}\frac{\partial\underline{E}}{\partial t} + \frac{i}{c}\nabla\times\underline{E} + \frac{i}{c}\frac{\partial\underline{B}}{\partial t}$$

;now, $\nabla\cdot\underline{E}/c = \rho/\epsilon_0 c = \mu_0 c\rho/\mu_0\epsilon_0 c^2 = \mu_0 c\rho$. Also $\nabla\times\underline{E} = -\partial\underline{B}/\partial t$ and $\nabla\times H - \partial D/\partial t = J$ or $\nabla\times\underline{B} - \mu_0\epsilon_0\partial\underline{E}/\partial t = \mu_0 J$. Therefore, $\Box A = \partial^{\dagger}\partial A = \mu_0 J$

[If we had used minquats for our basis we would have obtained:

$$\partial^{\dagger}A = \underline{B} - i\underline{E}/c = -i(\underline{E}/c + i\underline{B}) \quad \text{and} \quad \Box^2 A = -\mu_0 J.$$]

We also note that $\psi^*\psi = (-\underline{E}/c - i\underline{B})(-\underline{E}/c + i\underline{B}) = E^2/c^2 + B^2 + 2\vec{E}/c \times \vec{B} = -2\mu T^{\mu\nu}\sigma_\nu = 2\mu_0$(energy density + Poynting vector/c).

If we perform a Lorentz transformation on ψ we must use $L^{-1}\psi L = \psi'$ because ψ is an axial 4-vector(it can be considered as the hyper-curl of A). We then have: $L^{\dagger}\psi L =$ (for a boost in the z direction)

$$(\cosh\tfrac{1}{2}\alpha + \sigma_3\sinh\tfrac{1}{2}\alpha)(-\underline{E} + i\underline{B})(\cosh\tfrac{1}{2}\alpha - \sigma_3\sinh\tfrac{1}{2}\alpha) =$$

$$\cosh^2\tfrac{\alpha}{2}\left[(-E_1+iB_1)\sigma_1 + (-E_2+iB_2)\sigma_2 + (-E_3+iB_3)\sigma_3\right] + \cosh\tfrac{\alpha}{2}\sinh\tfrac{\alpha}{2}\Big[(-E_1+iB_1)(+i\sigma_2) + (E_2+iB_2)(-i\sigma_1)$$
$$+ (-E_3+iB_3)\Big] - \cosh\tfrac{\alpha}{2}\sinh\tfrac{\alpha}{2}\left[(-E_1+iB_1)(-i\sigma_2) + (-E_2+iB_2)(i\sigma_1) + (-E_3+iB_3)(i)\right] +$$
$$\sinh^2\tfrac{\alpha}{2}\left[(-E_1+iB_1)(+\sigma_1) + (-E_2+iB_2)(+\sigma_2) - (-E_3+iB_3)\sigma_3\right], \text{ so}$$

$$(-\underline{E}'+i\underline{B}') = (-E_3+iB_3)\sigma_3 + \cosh\alpha(-E_1+iB_1)\sigma_1 + \cosh\alpha(-E_2+iB_2)\sigma_2 - \sinh\alpha\,\sigma_1 i(-E_2+iB_2)$$
$$+ \sinh\alpha\,\sigma_2 i(-E_1+iB_1).$$ Comparing, we see:

$$\boxed{\begin{array}{ll} B_3' = B_3 & E_3' = E_3 \\ B_1' = \gamma(B_1+\beta E_2) & E_1' = \gamma(E_1-\beta B_2) \\ B_2' = \gamma(B_2-\beta E_1) & E_2' = \gamma(E_2+\beta B_1) \\ B_\perp' = \gamma\left(B_\perp - \frac{v\times E}{c^2}\right) & E_\perp' = \gamma(E_\perp + \vec{v}\times\vec{B}) \end{array}}$$

or $\underline{E_\parallel' = E_\parallel}$, $\underline{B_\parallel' = B_\parallel}$

Notice that the explicit use of the electromagnetic field tensor was not required.

LIE GROUP REPRESENTATIONS

DAVE PETERSON

Abstract. The popular continuous Lie Groups used for particle physics have their own representation conventions that differ from those used by mathematicians. The following note focuses on mathematical physics in preference over pure mathematics. The discussion here elaborates beyond a previous note on the Lie group SU(2) [4]. A goal has been to understand the decuplet and octet structure of baryon groupings in $SU(3)_F$ as indicated in equation (4) shown below. Actually, the decuplet **10** can be constructed intuitively since all quark spins are aligned and the states are completely favor symmetric. But the other multiplet require fairly advanced mathematics. Part of the new thinking is building on what we know about angular momentum $|J,m\rangle$ and applying it to isotopic spin $|I,I_3\rangle$. [Preliminary].

1. Introduction

Prior to Heisenberg's quantum mechanics of 1927, few physicists knew anything about matrices. And few physicists knew much about group theory until the power of of Gell-Mann's approximate Lie group SU(3) was revealed in 1964 for the baryons. The mathematicians knew, and Gell-Mann could have saved much effort if he had simply gone to a library and looked up their group classification schemes [2]. Lie groups go back to the studies of Sophus Lie who published them near the year 1890. Largely because of particle physics, Lie groups are now very popular. But it is still a difficult study.

In texts on particle physics, the irreducible representations of useful Lie Groups are often labeled in a simple form such as: $\mathbf{2}\otimes\mathbf{2}=\mathbf{3}\oplus\mathbf{1}$
where the numbers refer to the sizes or dimensions of multiplets. This example can pertain to the special unitary group SU(2), and we wish to extend this symbolism to higher groups such as SU(3). What does this labeling mean? and How can it help understand the particle multiplets of high energy physics?

There is a desire to be able to consider difficult concepts from perhaps a second year of graduate school and try to explain them at a sophomore level of college physics. This is not always possible, but here we see that the concept of the baryon decuplet **can** be explained simply. I am not aware that this simplicity is presented in any popular book, so I do so here.

The known baryons of the 1960's only consisted of three basic quarks, $\{u,\ d,\ s\}$ (now

Date: June 7, 2015.

email: davepeterson137@gmail.com. Paper updated to June 26, 2015.

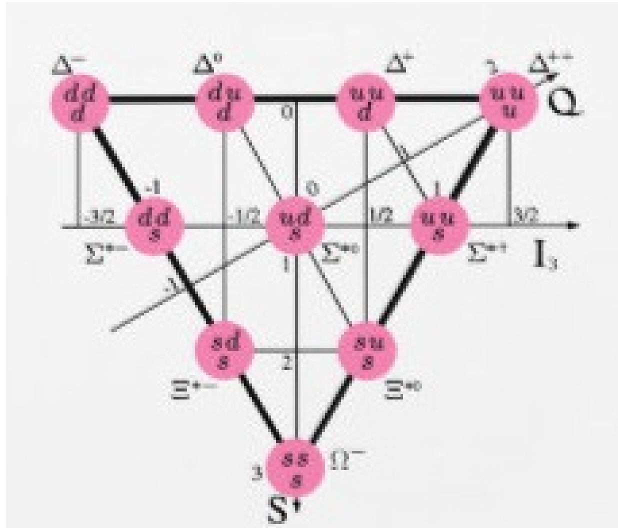


FIGURE 1. Plot of Baryon Decuplet, S versus I_3, showing the name of the particles and their triplet quark contents.

called "flavors"). Laying out all the particles in terms of how many of each quark might require 3-dimensions (s, u, and d-axes). Indeed, a convention is to show s'ness on a vertical y-axis. But a Heisenberg-Gell-Mann convention introduced an x-axis of "u-ness minus d-ness" thus enabling a **2-D** plot of known particles (and this x-axis was given the name "projection of isotopic spin" or I_3). A layout then includes particles like ddd, uuu, and sss (which are "corner" states of a triangular decuplet of particles- see Figure 1). Having something like uuu as a fermion required knowledge of "color" (red, green, blue from $SU(3)_{color}$). So uuu was allowed because it was really a red-u, green-u, and blue-u (a particle called the Δ^{++}). And uuu is also three spins up $|\uparrow\uparrow\uparrow\rangle$, so any group including uuu would be a net 3/2's spin state. The uuu state has the simplest possible symmetric wave functions, so it is easy to write out the states of the decuplet. At the time that Gell-Mann presented his plot, the sss state had not yet been detected (now called the Omega-minus particle).This and the more difficult multiplets are elaborated below. A fuller understanding requires knowing Lie groups for physics and might require going through courses with hundreds of pages of text and problems, e.g., [7].

First, some preliminaries:

2. Definitions

Group: A group, G, is a set of elements and an operation that composes any two elements into a third element belonging to the group. It has to obey this closure, be associative, have an identity element (e), and each element has to be invertible so that for all $g \in G$, $g^{-1}g = gg^{-1} = e$, the identity. Examples of groups include the integers (Z), the rationals (Q), and symmetric groups. But here, we care about Lie groups which are smooth continuous groups that are locally Euclidean (differential manifolds) such as the rotation groups like $O(2)\,and\,O(3)$.

General Linear Group: A convenient reference group representation is the set of square invertible $n \times n$ matrices such as **GL(n,R)** over the reals (determinant $\neq 0$) or GL(n,C) over complex numbers. The group concept applies because the product of two invertible matrices is also invertible, and the inverse of an invertible matrix is invertible [3]. The word linear applies because the columns of an invertible matrix are linearly independent. These are Lie groups of dimension $= n^2$.

Group Representation: Representing a group, G, by a square matrices, D(g), is convenient because the group operation can be represented by standard matrix multiplications from well-understood linear algebra. In physics, group symmetry can then be exploited for calculations. A representation is then a map over some vector space, V, preserving the group operation:

$D\colon G \to \mathrm{GL}(V)$ such that $D(g_1 g_2) = D(g_1) D(g_2), \qquad$ for all $g_1, g_2 \in G$.

A simple example is rotation in the plane by 120^o or $u = e^{2\pi i/3}$. Then the discrete cyclic group of successive transformations

$$C3 = \{1, u, u^2\} \qquad D\left(1\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad D\left(u\right) = \begin{bmatrix} 1 & 0 \\ 0 & u \end{bmatrix} \qquad D\left(u^2\right) = \begin{bmatrix} 1 & 0 \\ 0 & u^2 \end{bmatrix}. \tag{1}$$

A representation $T : G \to GL(V)$ is called **reducible** if there are subspaces W and Y such that $V = W \oplus Y$ where both W and Y are invariant under all the $T_g{}'s$. If no such subspaces exist, then V is **irreducible**.

Tensors: Instead of matrices, we can consider tensors as furnishing representations of a group. Tensors are objects that transform as if they are the products of vectors. In older notation, "dyadics" are second order tensors formed by juxtaposing pairs of vectors like the tensor product of vectors, $\vec{a} \otimes \vec{b}$.

A tensor of rank r is the direct product of r copies of a fundamental representation space (e.g., 3-vectors for SO(3) and 2-spinors for SU(2))[2]. Tensor products are often reducible and need to be decomposed into irreducible components (irreps) for similar physical properties (for example sets of symmetric versus anti-symmetric states). Considering tensors as representations also requires being careful with subscripts (in the basis of the defining representation) and superscripts (contravariant or conjugate representation). For example, if $\{e_1,\, e_2,\, e_3\} = \{u,\, d,\, s\}$ quark flavors, then anti-quarks may be $\{\bar{u},\, \bar{d},\, \bar{s}\} = \{e^1,\, e^2,\, e^3\}$. These may be labeled by bold numbers: $\mathbf{3}\,or\,\bar{\mathbf{3}}$. Tensor products form new bases such as $e_{11} = e_1 \otimes e_1,\; e_{ij} = e_i \otimes e_j$. Vectors over bases can be defined: e.g., $v = v^i e_i\,or\,\nu = v_i{}^j e^i e_j$.

Unitary: A complex square matrix U is said to be unitary if $U^\dagger U = UU^\dagger = I$ where I is the identity matrix and $U^\dagger$ is the conjugate transpose of U. A "special unitary group," SU(n) of degree n is the group of $n \times n$ unitary matrices with determinant 1 so that

$SU(n) \subset U(n) \subset GL(n, C)$. These groups are key to the standard model of particle physics: $\mathbf{G_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y}$ where the subscripts C means color, L means left, and Y means weak-hypercharge. The product of Lie groups is also a Lie group. Unlike $SU(3)_F$, $SU(3)_C$ has exact color symmetry.

Orthogonal Group: O(n) is the group of distance-preserving transformations of a Euclidean space E^n while preserving a fixed point in that space. The group operation is composition of transformations. As $n \times n$ matrices, the determinant has to be ± 1. The term orthogonal means that the inverse is equal to the transpose of the matrix: $Q^T Q = QQ^T = I$. The "special orthogonal group," SO(n) has $det = +1$ and is also called a rotation group about a fixed point (such as SO(2) or SO(3)). For physics, the angular momentum operators $\{J_x,\ J_y,\ J_z\}$ can be generators of SO(3).

Lie Algebra: Physicists refer to the Lie algebra as the space of infinitesimal Lie group elements. For SU(n), the Lie algebra is denoted by lower case su(n) as the set of traceless hermitian $n \times n$ complex matrices (and having a Lie Bracket given by $-i$ times the commutator). For example, let $g_z(\phi_1)$ be the first Euler rotation about the z axis in the space E^3 by an angle ϕ_1 in the continuous 3-D rotation group $O(3)$ [4]. Then,

$$g_3 = \left[\frac{dg_z(\phi_1)}{d\phi_1}\right]_o = \frac{d}{d\phi_1}\begin{pmatrix} \cos\phi_1 & \sin\phi_1 & 0 \\ -\sin\phi_1 & \cos\phi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\phi_1 = o} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2}$$

where the derivative applies on each element of the matrix. Finding these 'tangent matrices' takes the Lie Group to the Lie Algebra.

One can also go <u>backwards</u> from the basic generator g_3 to the general group element rotation $\mathbf{R}$ in terms of cosines and sines by evaluating the <u>'exponential map'</u>:
$\mathbf{R_3 = exp(g_3}\,\varphi) \equiv I + g_3\varphi + (g_3\varphi)^2/2! + ...$ where a matrix squared means the matrix times the matrix. Notice that the upper-left 2×2 sub-rotation of g_3 (last term of equation (2)) happens to be the quaternion $q_y = i\sigma_y$ for SU(2).

The number of elements in a square matrix or tensor, T, is n^2. This can be decomposed into symmetric and anti-symmetric (or skew) elements according to:

$$S^{ij} = \tfrac{1}{2}(T^{ij} + T^{ji}),\ A^{ij} = \tfrac{1}{2}(T^{ij} - T^{ji}),\ n^2 = \tfrac{1}{2}(n^2 + n + n^2 - n) = \tfrac{1}{2}n(n+1) + \tfrac{1}{2}n(n-1). \tag{3}$$

So, for n = 2, $\mathbf{2 \otimes 2 = 3 \oplus 1}$ for symmetric plus skew parts (as mentioned in the introduction above). A particular example [1] is the addition of two spin $\frac{1}{2}$ particles with spin states $\uparrow$ & $\downarrow$ where the symmetric total spin one states have spin projections $M_s = 1, 0, \& -1$ (a triplet) while total spin zero is a single anti-symmetric state:
$|S = 0, M_s = 0\rangle = \sqrt{\tfrac{1}{2}}\,(\uparrow\downarrow - \downarrow\uparrow)$.

And for n = 3, $3 \otimes 3 = 6 \oplus 3$ or sometimes written as $6 \oplus \bar{3}$, where the bar emphasizes anti-symmetry under the exchange of two given particles. Some use the over-bar for anti-particles so that a 3 representation for u,d, and s quark flavors would have $\bar{3} = \{\bar{u},\ \bar{d},\ \bar{s}\}$. If we are instead working with QCD colors, then $3 = \{r,\ g,\ b\}$ and $\bar{3} = \{\bar{r},\ \bar{g},\ \bar{b}\}$ anti-colors.

The group SU(3) has 3×3 unitary, unimodular matrices whose 8 independent generators are usually chosen to be the 8 Gell-Mann matrices, λ_i. This is similar to the use of the Pauli matrices $\{\sigma_x,\ \sigma_y,\ \sigma_z\}$ for SU(2).

One often wishes to treat the trace of a tensor separately so that a tensor product of two vectors is decomposed as the addition of a traceless-symmetric part + the trace + the anti-symmetric part. Then the above example for n = 2 becomes $\mathbf{2} \otimes \mathbf{2} = \mathbf{2} \oplus \mathbf{1} \oplus \mathbf{1}$
And for n = 3, $\mathbf{3} \otimes \mathbf{3} = \mathbf{5} \oplus \mathbf{1} \oplus \mathbf{3}$ More complicated tensor products are often treated using something called the “Young Tableaux.”

3. Flavor Multiplets:

“The history of nuclear and particle physics is very much a quest to find symmetry groups” [2]. As a generalization of Heisenberg’s original isospin SU(2) group for $\{p,\ n\}$, Gell-Mann came up with an SU(3) group for baryons. We would now call this a quark “flavor” group. $SU(3)_F$ is only approximate, and its utility was an accident of history due to the lightest quarks having a mass much below the mass of baryons.

One of the best known patterns built up from u, d, s quark constituents is the **“baryon decuplet”** [5] with each baryon containing three “valence” quarks. Strangeness (or hyper charge $Y = B + S$) is plotted versus Isotopic spin ($S\ vs.\ I_3$) leading to the prediction of a previously unknown “Omega minus” $\Omega^-(sss)$ particle of strangeness $S = -3$ (see Fig. 1). The **10** plotted baryon states of lowest mass have spin-parity $J^P = \frac{3}{2}^+$ and include the Δ particles, the Σ’s, the Ξ’s (“Cascade” particles) and the Ω^-. The wave functions of all these states is symmetric under interchange of any pair of quarks [5]. That means that all possible superpositions have only positive additions (+). Unlike the other multiplets, this makes the decuplet fairly easy to understand intuitively with a minimum of math.

The total number of states formed from $\{u,\ d\ ,s\ \}$ is $3 \times 3 \times 3 = 27$. Of these, one state is antisymmetric, $dsu + uds + sud - usd - sdu - dus$, leaving 16 states in two octets having mixed symmetry. The one containing the proton and neutron has spin-parity $J^P = \frac{1}{2}^+$. This octet also contains Σ’s and Ξ’s of lower mass-energy than those in the decuplet. And in the middle is the famous Lambda $\Lambda(sud)$ particle. Some of these wave functions are very complex with many carefully placed minus signs. The proton, for example, now has 12 terms still preserving an overall positive symmetry. Overall, the matter baryons are symbolized by [1]:

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = (\mathbf{3} \otimes \mathbf{3}) \otimes \mathbf{3} = \mathbf{27} = (\mathbf{6} \oplus \mathbf{3}) \otimes \mathbf{3} = (\mathbf{6} \otimes \mathbf{3}) \oplus (\mathbf{3} \otimes \mathbf{3}) = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \tag{4}$$

[1] Yes, $6 \otimes 3 = 10 \oplus 8$ can be shown using the Young Tableaux (e.g., Palash Pal, An Introductory Course of Particle Physics, CRC, 2015, HW p 267).

The mesons are characterized by a quark and an anti-quark, and these can have aligned spins ($J = 1$) or opposite spins ($J = 0$) forming two different nonet meson groupings. For mesons, baryon number $B = 0$ and spin-parity $J^P = 0^-$ are called pseudoscalars mesons and form an nonet of particles while $J^P = 1^-$ form a nonet of vector mesons [5] with the same quark assignments for both nonets. Fermions (like quarks) and antifermions have opposite intrinsic parity. Note that $3^2 = 9$, or $3 \otimes \bar{3} = 8 \oplus 1$. The neutral pion is in an isospin triplet with the $\pi^\pm$'s and has wave function $\pi^o : \frac{1}{\sqrt{2}}(d\bar{d} - u\bar{u})$. The combination $\frac{1}{\sqrt{2}}(d\bar{d} + u\bar{u})$ is an isospin singlet state called the eta η (550 MeV) meson.

4. Simple Approach to Baryon Multiplets:

For the strong interactions, in 1932 Werner Heisenberg proposed that the proton and neutron be considered as nearly the same particle (a nucleon) but with a kind of spin-up and a spin-down difference. In 1937, Eugene Wigner gave this concept the name Isospin, I, a value that is preserved under strong interactions. The number of particles in a similar mass multiplet is $n = 2I+1 = 2(\frac{1}{2})+1 = 2$ for this doublet, and spin projection $I_{3\,p} = +1/2$ and $I_{3\,n} = -1/2$. When pi-mesons were discovered in 1947, they were considered as an isosopin-triplet: $I = 1$, $I_3(\pi^+) = +1$, $I_3(\pi^0) = 0$, $I_3(\pi^-) = -1$. It was then discovered that the baryons were composed of three quarks, so the baryon number of the quarks had to be $B = 1/3$. Since a proton is $p = \{uud\}$ and a neutron is $n = \{udd\}$, the original isospin projection, I_3, became a measure of "u-ness:" $I_3(u) = +1/2$ and $I_3(d) = -1/2$. Then with the charge formula: $Q/e = \frac{1}{2}(B + S) + I_3$, we see that the charge of the u-quark must be $Q = +2/3$ and the charge of the d-quark is $Q = -1/3$. The strange quark has strangeness $S = -1$ (an unfortunate historical convention like Ben Franklin calling the electron charge negative). It has no isotopic spin, so its charge is $Q = -1/3$.

With this basic (or fundamental) understanding, we can begin to plot out quark-triplet combinations on a chart with net isotopic spin on an x-axis and strangeness on a y-axis and consider only the selections of quarks: u, d, and s. A top row of 4 particles with no strangeness ($S = 0$), will end with a ddd triplet on the left corner and a uuu triplet on the right corner. It turns out that these corners correspond to particles called the Δ^- ($I = 3/2, I_3 = -3/2$) and Δ^{++} ($I_3 = +3/2$) [2] . Quarks are fermions, so these triplets can only happen if each quark is different somehow. The somehow is having different colors: r, g, and b. Progressing down on the chart, we have another triangular corner at sss with strangeness $S = -3$ ($I = 0$, $I_3 = 0$). This is called the Ω^- particle (see Fig. 1).

The total number of particles in this overall triangle is $n = 10$. Now the total number of combinations for triplets is $3 \times 3 \times 3 = 27$, so these 10 must have some property that makes them fit into a special multiplet. Since the corner states are symmetric under interchange of the quark order in the triplet [5], that property must be <u>symmetry</u> of the wave-function of the particle states (flavor symmetry). So, if we include the corner

[2] Δ^{++} is called a maximum weight state, and other states can be formed from it by applying lowering operators.

states, we must have a symmetric decuplet, $n = 10$ particles. We can label this group as $J^P = \frac{3}{2}^+$ with all spins aligned (e.g., $\{\uparrow\uparrow\uparrow\}$). An example of states is $\psi(\Delta^-) = |ddd\rangle$, and $\psi(\Delta^+) = |uud + udu + duu\rangle/\sqrt{3}$. All of these states are simply composed of all combinations of the quarks for a given state. This makes the decuplet simple and intuitive. But all other multiplets are more complex and require some sophisticated mathematics. All these triplet particles are fermions, so the net anti-symmetry comes from antisymmetry of the color terms.

If we don't include these corner states (delete them from the chart), then the size of the multiplet is reduced (into octets). We no longer have all spins aligned, so we have to consider terms like $|\uparrow\uparrow\downarrow\rangle$ and $J = 1/2$. One of the baryon octets includes the proton and neutron and is also symmetric $J^P = \frac{1}{2}^+$. But it has mixed spin and flavor and the overall spin-flavor symmetry is achieved by cyclic permutations (which is lengthy and complex, the proton wave function now has 9 terms in it to achieve this symmetry).

The mesons form two groupings called $J^P = 0^-$ for pseudo-scalar mesons [3] and $J^P = 1^-$ for vector-mesons. For example, both the π^o and the ρ^o have $|u\bar{u} - d\bar{d}\rangle/\sqrt{2}$, but the π^o has anti-aligned spins while the ρ^o has aligned spins for the quark and anti-quark pairs. The plotting of the mesons by S versus I_3 is straightforward. One interesting different way to do it is by Martinus Veltman [6] where he starts with an inverted triangle for the fundamental $\{sud\}$ representation and then adds three conjugate triangles $\{\bar{s}, \bar{u}, \bar{d}\}$ pointed up and centered at each vertex of the initial triangle. The vertices of each triangle represent a type of antiquark and antiquark combination (kaons and pion totaling 6 on the outside). The center has three particles (pi-zero and eta mesons). But, the details of the wave functions can be tricky (require Clebsch-Gordon coefficients or some other advanced method).

References

[1] Francis Halzen, Alan D. Martin, Quarks & Leptons, An Introductory Course in Modern Particle Physics, John Wiley & Sons, 1984.

[2] Notes by Prof. Hugh Osborn, Part I: "Symmetries and particles," University of Cambridge, Department of Applied Mathematics and Theoretical Physics (DAMTP). http://www.damtp.cam.ac.uk/user/examples/3P2Lb.pdf .

[3] General Linear group, http://en.wikipedia.org/wiki/General_linear_group .

[4] Dave Peterson, "Electron spin and SU(2)," 10/20/12, (in A Stroll Through Physics, DP, March, 2014 (Spin.tex).

[5] Donald H. Perkins, Introduction to High Energy Physics, 4th Edition, Cambridge, 2000.

[6] Martinus Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, 2003, p. 273.

[7] E.g., 258 pages course. Berkely Lectures http://math.northwestern.edu/∼ theojf/ LieQuantumGroups.pdf Or Joel A. Shapiro, "Group Theory in Physics," January 16, 2012

[3] A pseudoscalar is a quantity that behaves like a scalar but changes sign under a parity inversion while a true scalar does not. A standard example is the product $\vec{a} \cdot (\vec{b} \times \vec{c})$ which changes sign under mirror reflection.

https://www.physics. rutgers.edu/grad/618/lects/lects_2.pdf 158 Pages. Or: http://www.hep.man.ac.uk/u/pilaftsi/SYM/sym.pdf .

Geometry in Modern Physics

Dave Peterson, 4/2/16 - 7/16/16, (Preliminary Revision_0)

The following is a sketch of topology and geometry needed to better understand the increasingly popular but advanced works of **Shiing-Shen Chern** (1911-2004). This summary background includes: differential forms, fibre-bundles, homotopy, homology, and topological invariants -- and books mentioning Chern usually discuss all this background first. A short introduction to Chern classes and Chern-Simons theory is then included. The motivation for this study is a desire to understand common applications of topology for modern physics and a possible mechanism for the production of cosmic matter asymmetry from SU(2) "sphalerons" at energies above the electro-weak symmetry breaking scale. In addition, modern topology has become a major player in the new exotic materials experiments of condensed matter physics (solid state theory and Bose-Einstein condensates.

The overall subject material is quite difficult. My approach to it has been to select the key defining statements from a variety of textbooks along with summary clarifications from a great many google references. This initial study is not yet "polished," and it may take a long time to do so.

Simplistically, a goal of topology is to present enough topological invariants (usually integers) to enable characterization of "spaces." Initially, this often involves specifying n-dimensional "holes" in spaces. But in what sense does a physical field represented by complex groups such as SU(2) have holes?

One facet of the answer is that "holes" aren't restricted to our visual Euclidean concept of threading string about or through holes in familiar Euclidean space. Homology extends beyond these "1-holes" to "2-holes" (geometrical objects that can be filled with water – such as a 2-sphere or a torus). And there are "n-holes" in higher dimensions that we can't picture. Topology also includes complex and quaternionic spaces so that quantum mechanics can also be discussed. And there are also fields on manifolds and their associated connections and curvatures and forms and topological indices. And these fields may be discontinuous.

It may be that physics initially bumped into topology through "Dirac's monopole" in the 1930's, but "there is no doubt that a principal factor in the rise of topology in physics is due to the rise to supremacy of gauge theories in physics" in the 1970's [3]. The best known of these is the famous $SU(3)_c \times SU(2)_w \times U(1)_Y$. 't Hooft discussed a "non-Abelian" gauge theory monopole in 1972 and also made a major impact then by showing that such theories were renormalisable. This was followed by the importance of "instantons" in 1975.

Physics applications of Chern's differential geometry:

Shiing-Shen Chern (1911-2004) was a Chinese-born American mathematician and a major contributor to 20th century differential geometry. In his early years, he advanced mathematics in China and then later also worked in Berkeley from 1960 to 1979. In the 1940's, he co-authored studies with Weil on "Chern-Weil theory" for topological invariants of vector bundles. The concept of "Chern classes" was introduced in 1946 as topological characteristic classes largely using the language of forms. Chern-Simons theory ("**CS**") was produced in 1974 and inspired Edward Witten in some of his later contributions. The earlier Chern-Weil homomorphism was an important step in the theory of characteristic classes and is built into the construction of Chern-Simons forms.

Although intended mainly as pure mathematics, Chern-Simons forms were applied in physics for chiral anomalies, fractional statistics of anyons, electro-weak sphalerons, and instantons. Some of Chern's discoveries have earlier precedents that didn't catch on well at the time: Fiber space (then called sphere-space) was defined in 1935 by Hassler Whitney. Connections on fiber bundles were introduced in 1950 by C. Ehreshman – a student of Cartan. And Chern classes were inspired by Dirac's work several decades earlier. The algebra of differential forms goes back at least to 1899 with the publications of Élie Cartan. Physicists gained interest in topology from Yang-Mills theory [SU(2) for isospin]. Finite action of Yang-Mills theory is characterized by topological instanton number, Pontryagin index or second Chern Class. Some of this application is due to boundary conditions (BC's) on gauge potentials such as compactification of R^4 [20].

In mathematics and physics, a goal is to find and utilize invariants—properties that do not change under various transformations. Fundamental physics equations, for example, should be written in a form that has Lorentz invariance. General covariance gives forms that do not depend on any choice of coordinate system. Differential forms are also independent of coordinate choices.

For square matrices, A, eigenvectors and eigenvalues, determinant, and trace are invariant under changes of basis (e.g., beginning with the i, j, k, or e_i, e_j, e_k basis of Euclidean E^3). In quantum mechanics and linear algebra, students find eigenvalues and eigenvectors by using a **"characteristic polynomial"** given by: $\mathbf{det(A - \lambda I) = 0}$ [and then solve the polynomial for its roots and λ values]. For fibre-bundles, connection independent information is given by constructing invariant polynomials in terms of "curvature forms, F " (e.g., the $F_{\mu\nu}$ of 4D electromagnetism).

The Chern modification of this looks something like this:

$$\det(it\,\mathbf{F}/2\pi + I) = \mathbf{\Sigma} c_k(V)t^k, \qquad \text{Eqn 1.}$$

where F is a curvature form of a vector bundle V, c_k are Chern forms, t is a variable, and I is the identity matrix (1's down the diagonal). Notice that if we select λ = -1 and replace the previous matrix A by a matrix of curvature 2-forms iF/2π , we can motivate the form of the Chern polynomial from the characteristic polynomial.

The "action" S of Chern-Simons theory is proportional to the integral of the CS 3-form, c_3: $S = (k/4\pi) \int \mathrm{tr}(A\wedge \mathbf{d}A + (^2/_3) A\wedge A\wedge A)$. Eqn 2.
{exterior product Λ discussed further later. And most of Physics can be described using it}. Ed Witten used this CS action to obtain the Jones polynomial of knot theory.

Background:

DIFFERENTIAL FORMS:

Differential forms are covered under the topic of "exterior calculus" which enables abstraction without reference to any specific coordinate system [23]. Flanders [16] says that this "exterior calculus is here to stay and will gradually replace tensor methods in numerous situations where it is the more natural tool."

In a loose sense, differential forms are integrands (what's under the integral sign) over one-dimensional curves, 2-d surface areas, 3-d volumes, or higher n-d manifolds. They intend to be independent of particular coordinates and possess orientations. Their utility in modern topology and geometry was pioneered by Cartan around 1900. For example, an integral of a function $\int_a^b f(x)dx$ motivates the idea of a one-dimensional "1-form," $f(x)dx$, with an orientation given by the limits from a to b. But forms go deeper than that. It is wrong to think of dx as a "tiny Δ" in this arena but something that rather acts more like a "basis" vector. A form is like a "functional" machine acting on a coordinate vector in tangent vector space. Then, **d**x as a form acts on a tangent vector v_p to give a coordinate value of v at point p. In linear algebra, 1-forms are naturally "dual" to vector fields on a manifold. Integration of differential forms is only well-defined on oriented manifolds (spaces that are locally like flat Euclidean space).

Forms are often expressed using "**wedge** products" (or "**exterior** or **alternating** products") such as bivector $a = u \wedge v$, and the wedge product of two 1-forms is a 2-form. In Euclidean 3-space, $u \wedge v \simeq u \times v$ (the familiar vector cross-product), and both have an orientation and represent the area of a parallelogram formed by the u,v vector sides. Just as $u\times v = -v\times u$, $u \wedge v = -v \wedge u$, and $u \wedge u = -u \wedge u = 0$ (a parallelogram with no area). But, 3-d cross-products cease to apply to 4-d space, and there $u \wedge u$ may not be zero—especially if u is a 2-form.

"Wedge" is **defined** as: $\mathbf{u} \wedge \mathbf{v} = \Sigma_{i<j} (u_i v_j - u_j v_i)\, e^i \wedge e^j$, where e's are basis.

This is also called: $u \wedge v = u \otimes v - v \otimes u$ (**skew symmetric**,tensor product).

For R^3, the index values i< j only allow (i,j) = (1,2),(1,3),(2,3), and $e^1 \wedge e^2 \sim e^3$ (cyclic). Expanding the definition gives the same result as $u\times v$, cross-product.
Wedge products distribute: $u \wedge (v+w) = u \wedge v + u \wedge w$. And they are associative: $u \wedge (v \wedge w) = (u \wedge v) \wedge w$. We often write a wedge product of $dx \wedge dy$ as just dxdy.
One may also wedge product two **matrix-valued forms** such as A = (a,b; c,d) , a 2× 2 matrix, and can then calculate $A \wedge A$.
"If x^i are coordinates, then dx^i are a local basis for 1-forms" [17]. Exterior products generate local bases for higher order forms, e.g., $dx^1 \wedge dx^2$.
A symbol for the set of all p-forms is $\mathbf{\Lambda}^p$.

When he was 32, Cartan also created the concept of an **"exterior derivative" d** which takes a 1-form into a 2-form or an n-form to an n+1 form. An example of a 2-form in E^3 is the magnetic field, B. Although we often call B a vector, it really flips under mirror image as a pseudo-vector.
Using the shortened notation, $\partial_x = \partial / \partial x$, the operator $\mathbf{d} = [\partial_x dx + \partial_y dy + \partial_z dz] \wedge$ ..
$B = \nabla \times A = \Sigma(\partial_i A_j)\, dx^i \wedge dx^j = \mathbf{d}A$.
We usually write the vector potential using unit vectors as $A = iA_x + jA_y + kA_z$, but we can also write it as a 1-form: $A = A_1 dx^1 + A_2 dx^2 + A_3 dx^3$ (superscripts here are not "powers" but simply index numbers). It should be a 1-form like momentum or wavenumber because A is a kind of electromagnetic momentum. Calculating the vector sum or the form sum gives the same answers (with the understanding that unit vector k is now $dx^1 \wedge dx^2$— showing it more clearly as a 2-form.

In Euclidean space $\mathbf{E^3}$, vector differential operators have a simple form:

If f is a 0-form (a function), A a 1-form, and B a 2-form, then grad f = $\nabla f \rightarrow \mathbf{d}f$. Curl A = $\nabla \times A \rightarrow \mathbf{d}A$, and div B = $\nabla \cdot B \rightarrow \mathbf{d}B$. "Poincare's lemma" [23] says that if a form V obeys $\mathbf{d}V = 0$, then V is closed; and if $V = \mathbf{d}U$, then V is "exact." Then $\mathbf{dd}U = 0$.
The operations curl grad f = 0 $\equiv$ $\mathbf{dd}f = 0$; and div curl A = 0 $\equiv$ $\mathbf{dd}A = 0$. These null operations are the ones enabling the existence of electromagnetic potentials ϕ or f and vector A. A few basic rules for application of the exterior derivative are:
$\mathbf{d(df)} = 0$, $\mathbf{d}(\alpha \wedge \beta) = \mathbf{d}\alpha \wedge \beta + (\alpha \wedge \mathbf{d}\beta)(-1)^p$, $\mathbf{d}(f \wedge \alpha) = \mathbf{d}f \wedge \alpha + f \wedge \mathbf{d}\alpha$.

If A is a 1-form electromagnetic **4**-vector potential (say A^1, A^2,A^3, A^4— or A^0), then we can take F = **d**A as an anti-symmetric 2-form (where **F** [standing for "**Faraday**"] is now related to the electromagnetic tensor $F_{\mu\nu}$, and **d** might be called a "generalized curl" operator). {Considering $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ requires accounting for the metric $\eta_{\mu\nu}$ sign (note: $\partial_\mu \equiv \partial / \partial x^\mu$ and A_μ is covariant – more later.)}. The Faraday 2-form is sometimes labeled as $\omega = F_{\mu\nu}\, dx^\mu \wedge dx^\nu = 2F$ (index duplication doubling). So **F** = ω/2, or sum Σ restrictions μ < ν to avoid this duplication.
One may also write the 2-form F in terms of electromagnetic fields as:

$$\mathbf{F} = (E_1dx+E_2dy+E_3dz) \wedge dt + B_1dy \wedge dz + B_2dz \wedge dx + B_3dx \wedge dy.$$

(note that the signs of $Edt \wedge dx$ depend on metric sign: - +++ gives –E).
Cutting back from 4-space to 3-space, F pulls back to the magnetic field 2-form B as: $\mathbf{F}_{|R3} = B$ as a curvature on R^3.

Saying that **dd**A = **d**F = 0 only helps to characterize the "homogeneous" Maxwell equations. For the rest we have to say d*F = J using a new concept called the "Hodge-star, *, and a 3-form current source J. "Star" is a mapping to complementary dimensions: $*:\Lambda^p \rightarrow \Lambda^{n-p}$ [Burke]. For example, a 2-form in E^3 such as α = pdydz + qdzdx + rdxdy has *α = pdx + qdy + rdz. That is, calculus can be extended to n-d metric spaces using the Hodge star.
For the 3-space discussion about the magnetic field above, saying that $dx \wedge dy$ is like k is stated more appropriately using star: $*dz = dx \wedge dy$ or $*dx \wedge dy = dz$. More careful presentations on electromagnetism use 1-forms for E and H but 2-forms for D and B and current J and charge density ρxdydz is a 3-form.

In 4-dimensional relativity, momentum is a 4-vector p^μ (~tangent vector); or with sub-scripts, p_μ is a differential 1-form [ref. MTW]. Mathematics books often don't pay much attention to whether a superscript or subscript is used. But physicists care and strive to note the difference between using superscripts for "contravariant" vectors and subscripts for "covariant" vectors.
Momentum and wavenumber, k, are more properly covariant (k_μ) or 1-forms that can be thought of as similar to a picture of flat equally spaced surfaces. A dot product u·v is also $u_\mu v^\mu = u \cdot v = \eta_{\mu\nu} u^\mu v^\nu$ (where repeated indices mean sum through all their values—the "Einstein convention," and eta refers to the special relativity Lorentz metric—diagonal ± 1's). Momentum in 4-d as a 1-form is $p = -Edt + p_xdx + p_ydy + p_zdz$.

Dual Space:

Given any vector space V with any basis, its "dual space" V* is a set of "linear functionals" of form $\phi(x,y) = ax + by$. For example, suppose we have $v^1=(1,0)$ & $v^2 =$

(1,1) in R^2. Then the dual basis is $\phi_1(x,y) = x-y$, $\phi_2 = y$] such that $\phi_i(v^j) = \delta_i^{\,j}$. {This is like the usual basis statement that the dual basis is defined as $\mathbf{e}_i(\mathbf{e}^j) = \delta_i^j$ }.
{Verify: $\phi_1(v^1) = a_1 = \phi_2(v^2) = a_2+b_2 = 1$; and $\phi_1(v^2) = a_1+b_1 = 0 = \phi_2(v^1) = a_2$, check.}.
Picking any vector or point p = (x,y) in R^2 gives number values to the $\phi_i(x,y)$'s as weighing coefficients of the vector. Then vector $v = (x,y) = \phi_1(x,y)\, \mathbf{v^1} + \phi_2(x,y)\, \mathbf{v^2}$.
Again using the example above, if $v = (x,y) = (3,2) = c_1v^1 + c_2v^2$, then $c_1 = 1$ and $c_2 = 2$. And indeed, $\phi_1(v) = x-y = 1$ and $\phi_2 = y = 2 = c_2$.

One may also think of dual bases as represented by row vectors consisting of the coefficients of dual ϕ's) and column vectors for coordinates. Then the inner product of $\langle\phi_1\, v^1\rangle$ is the row (1 -1) times the column(1 0) = 1.

Physicists are often taught vector analysis first with dot products as a special case of inner products and are later told that Dirac's "bra-ket" $\langle u|v\rangle$ is an inner product and that the "bra" is a dual vector. Examples include row vectors times column vectors to produce a number. It is almost like a scalar product between covariant and contravariant vectors. If v or the Dirac "ket" $|v\rangle$ is from a vector space, V, it is better to think of inner product bra-ket as a *"machine"* -- a bra vector is a functional that acts on a ket and spits out a number. And in linear algebra, a functional (also called a linear form or 1-form or "covector") is a linear map from a vector space to its field of scalars. In the previous paragraph, we have $\phi_i(v)$ as the action of a function of the vector to give a number (in this case a coefficient of the point in space). Or in bra-ket notation, $\phi \rightarrow \langle u|$, and $\phi_u(v) \rightarrow \langle u|v\rangle$.

Tangent Space:

For Euclidean space E^3, a tangent vector v_p consists of two points, "its vector part and a point of application" [O'Neill]. They can go off in any direction and have any length. $T_p(E^3)$ is the tangent space of E^3 at p. For any manifold M, its tangent space is the set of all tangent vectors at a point $p\in M$ over all points in manifold M and is labeled T_pM. A spherical surface like the 2-sphere S^2 has dimension two (e.g., over all θ and ϕ angles), and each point $p(\theta, \phi)\in S^2$ has a set of tangents to the surface over all directions and magnitudes. For any selected point on S^2, the tangent space is a plane tangent to the point p and also has dimension d=2. A tangent "bundle" TM over all points p then has "base" space as the sphere $B = S^2$ and fibers $F = T_pM$ and has total dimension dim[TM] = dim[B]+dim[F] = 2+2=4. Unlike ordinary vectors which can be positioned anywhere, a tangent vector is a vector attached to a particular point (only one allowed origin at a time). If we only consider unit tangents, their direction can be characterized simply by another angle $\alpha \in [0,2\pi)$. Then TM effectively has 3 dimensions θ,ϕ,α corresponding topologically to the rotation group SO(3). {Ref.[2], Frankel}.

All tangent vectors are called "contravariant" vectors. A vector field selects a particular v_p at each p. For a scalar function f and a vector v_p, there is a derivative of f with respect to v_p defined by: $\mathbf{v}_p[f] = (d/dt)(f(\mathbf{p}+t\mathbf{v}))|_{t=0}$, and this is called a **directional derivative.**

I like to write the first approximation to the "Taylor" series in the "pretty" form:

$$f(\mathbf{x}+\boldsymbol{\Delta}) \simeq f(\mathbf{x}) + \boldsymbol{\Delta}\cdot\boldsymbol{\nabla} f$$ (with $\boldsymbol{\Delta}$ being a small vector displacement).

The last term is called a "directional derivative," the projection of the gradient in the direction of the vector $\boldsymbol{\Delta}$. The "exterior derivative" takes the scalar field to a 1-form,

df, and the gradient is properly a 1-form. The component of ∇f in the direction Δ can then be written as $\langle \mathbf{\Delta}, \mathbf{d}f \rangle = \partial_{\Delta}(f) = \nabla_{\Delta} f$.

The gradient ∇f points towards the direction of the greatest rate of increase of the smooth scalar function f. $\nabla f = \mathbf{d}f = \mathbf{d}_{\wedge} f = (\partial f / \partial x^i)\, \mathbf{e}_i$.

$T_p(M)$ has a dual written as $T_p{}^*(M)$ consisting of one-forms and is called a "cotangent" space. A simple example of a cotangent bundle is taking gradient phi for every point of M.

<u>Covariant and Contravariant and transformations:</u>

"Vectors" are defined by the manner in which they <u>transform</u> from basis to basis. Note that it is rare for the words covariant and contravariant themselves to ever be defined. Wikipedia gives a one-dimensional (1D) example saying that if an axis is changed from units of meters to a smaller unit of centimeters, then the components of distance or velocity will be magnified by 100 (scales inversely or "<u>contra</u>"). In contrast, a gradient axis has units of 1/distance (as do dual vectors also called covectors). So coordinates of these "vectors" scale <u>with</u> the distance unit magnification – "co-". [The names came from James Joseph Sylvester in 1853]. By convention, contravariant vectors (or tangent vectors) use upper indices, like $\mathbf{v} = v^i e_i$ (repeated indices are summed over). And covectors use lower indices for their component values, $\mathbf{u} = u_i e^i$.

We say that contravariant components "transform <u>as</u> the coordinates do." For a coordinate transformation represented by a matrix, x' = Mx, a contravariant vector will also transform as v' = Mv. And a covariant vector components change oppositely to coordinates.

Another way to say this is that if we were to <u>change bases</u> in a vector space from coordinates (say) x^i to new coordinates y^i, then $x^i = \mathbf{\Sigma}\, (\partial x^i / \partial y^j)\, y^j$. Vectors that transform this way go by the name "contravariant." In contrast, a gradient of a scalar field, ∇f, transforms differently from this and is called a covariant vector with coordinates in a dual space. This difference in transformation becomes very important in tensor calculus and general relativity. One of the first sources to use the more abstract differential geometry and topology in physics was Misner and Wheeler in 1957.This was then developed further in their massive <u>**Gravitation**</u> book with author Kip Thorne ("**MTW**," 1973). A tensor is a multilinear map determined by its values on a basis and a dual basis (contravariant and covariant bases).

If a vector curve is parameterized as radial vector **r**(t) over components $x^i(t)$, then a tangent vector field is $\mathbf{T} = T^i = dx^i/dt$. Change coordinates to $y^i = y^i(x^1, x^2, \ldots x^n)$,$1 \leq i \leq n$. Then the tangent vector is $T' = T'^i = dy^i/dt = (dy^i/dx^j)(dx^j/dt) = T^j(dy^i/dx^j)$.

Frankel [2, pg 23,42]. A tangent vector is contravariant. Vectors are equivalent to their associated differential operator $\partial/\partial x^i \equiv \partial_i = e_i$. $v = \partial/\partial x^j\, v^j(x)$. That is, these basis vectors "corresponding to a coordinate system are tangent to the coordinate lines" motivate the notation $e_i = \partial/\partial x^i$. And "the coordinate basis one-forms are gradients of the coordinate surfaces," so $e^i = dx^i$. [It takes some time to get used to this way of thinking].

For intrinsic coordinate free concepts we consider a point p lying in a patch overlap of two open sets, $p \in \mathbf{U} \cap \mathbf{V}$, and transformations for each set are considered.

A 1-form is a covariant vector or convector and transforms as: $a^V{}_i = \Sigma a^U{}_j(\partial x_U{}^j / \partial x_V{}^i)$. But a contravariant vector $X^i{}_V = \Sigma\, (\partial x^i{}_V / \partial x^j{}_U) X_U{}^j$.

A tensor can be a mixture of covariant and contravariant vectors and may be expressed as: $\mathbf{T} = T^{ab}{}_{cd..} dx^c\, dx^d\, \partial_a \partial_b$. A goal of differential geometry and tensor forms is to be free of any one basis. That was a goal of general relativity too where choice of basis wasn't needed.

Definitions: Topology

Topology is a broad study that includes the disciplines of "point-set" topology, algebraic topology, and differential topology. General or point-set topology is the abstract study of the ideas of nearness and continuity (Wallace) and is an abstract foundation for "higher" studies in topology. Algebraic topology is the study of topological spaces and continuous functions using objects such as groups, rings, and homomorphisms. Differential topology is the study of those properties of a set which are invariant under diffeomorphisms (Milnor). And physicists probably care much more about differentiable manifolds than point-sets (with some interesting exceptions like Cantor sets). Felix Klein said that topology is the study of all properties of a space that are invariant under one-to-one bicontinuous mappings.

In topology and related areas of mathematics a topological **property** or topological invariant is a property of a topological space which is invariant under homeomorphisms [definition from Wikipedia]. Common examples include connectedness, dimension, compactness, "Hausdorffness," Euler characteristic, orientability, and algebraic invariants like homology, homotopy groups, and K-theory.

Today, the applications of topology in physics are numerous [27]; and the importance of topology emerged with the recognized importance of gauge theories. In this paper, we care mainly about applying topology to the field theories of modern physics. But the most obvious and productive applications have been in the realm of condensed matter (solid-state physics). A large aspect of this is the recent revolution called **"topological matter"** such as topological insulators, topological phases, topological superconductors and topological semimetals [27]. These topological phases are characterized by "topological invariants that have a global dependence on characteristic parameters of the system." Continuous deformation of one will not extend to those of another. Quantum effects are usually low energy modes and the topology of the bands of energy spectrum. The first big example of topological matter was the integer quantum Hall effect (IQHE).

As an example: crystals are characterized by repetitive arrangements of atoms or molecules and are easily pictured. Quantum mechanical wave functions there obey periodic boundary conditions such that $\psi(x+na) = \psi(x)$ where $n \in Z$ and a is a spacing between atoms. The electronic potential is also periodic this way, $V(x+na) = V(x)$. For a 2-d surface with atomic spacings a in the x direction and b in a y-direction, we can "identify" opposite sides of an $a \times b$ square on the surface. But this is also a description of a torus: fold up the longer opposite sides into a circle (joining the opposite edges) and then fold up the shorter opposite sides: $T^2 = S^1 \times S^1$. The symmetry of a crystalline solid defines what is called its "Brillouin torus." And over this torus one can define a "Bloch bundle" $E(T^2)$ [e.g., for the integer quantum hall effect (from 1981)].

More appropriately, a mathematical analysis of elastically coupled atoms in a linear periodic lattice [29,30] results in a “dispersion relation” $\omega = \omega(k) = A|\sin(ka/2)|$ with period $2\pi/a$ or a “first Brillouin zone”: $-\pi/a \le k \le +\pi/a$ where k is called “crystal momentum.” For 2d, there is a similar range for another k_2 in terms of $\pm\pi/b$. Then the effective torus is defined over this momentum space k_1 and k_2 rather than spatial x at a or b. Wavefunctions for these cases include Bragg reflections at these end points, both forward and reverse traveling waves, and opposite phases at the location of the atoms. The analysis automatically leads to the existence of forbidden gaps between allowed energy bands.

Homotopy and the Fundamental Group, $\pi_1(X)$:

Homotopy may be considered as the most important concept in topology. It is concerned with 1-D “loops” (say loop A and loop B) that might be continuously deformed into each other via a continuous map H. For 2-D spaces, we map from a square $[0,1]\times[0,1]$ with parameters t and s such that $H(t,0)$ is the loop $A(t)$ and $H(t,1)$ is the loop B. The first and simplest homotopy group is called the fundamental group, $\pi_1(X, x_o)$ [of a topological space X and a particular point in the space. Poincare, 1895]. In a simply connected space like R^n, all paths can be shrunk (contracted) to a point and $\pi_1 = 0$. For a circle, a path around the circle cannot be shrunk and it may loop around many times so that $\pi_1(S^1) = Z$ (the group of integers, Z for “Zahlen”). Also, $\pi_1(U(1)) = \pi_1(U(n)) = Z$ (not simply connected). For an object that is a product of two topological spaces, $\pi_1(X\times Y,(x_o,y_o)) \simeq \pi_1(X,x_o) \times \pi_1(Y,y_o)$.

So, as example, for a torus $T^2 = S^1 x\ S^1$, there are two classes of loops that cannot be shrunk and $\pi_1(T^2) = Z \oplus Z$ or $Z\times Z$. A cylinder is just $C = S^1\times I$, and a solid torus is $S^1\times D^2$; and π_1 for both is just Z. Homotopy doesn’t relate spaces themselves but rather an equivalent of homeomorphisms for functions between topological spaces. Perelman proved the Poincare conjecture that any 3D topological space X with $\pi_1(X)=0$ is topologically equivalent to the S^3 sphere [27].

The symbol π_o (S)=0 is used for the set of all path segments that can be deformed into each other. This is a simpler concept with just a one-way piece of string being able to connect any pair of points in space S. If a space is “simply connected,” then the symbols $\pi_o(S)=\pi_1(S)=0$.

“Straight-line homotopy” is the simplest example. Like a volume control knob, it progressively and linearly slides one curve into another with intermediate curve given by “gamma-path”: $\gamma_t(x) = (1-t)\gamma_0(x) + t\ \gamma_1(x)$ [as parameter t slides from value 0 to 1]. Any two loops in R^n are homotopic via this straight-line homotopy.

Many people already know that a coffee cup is topologically like a donut because they each have one hole and each could be continuously deformed into the other. And then a “deformation retract” of the donut reduces it to a circle (solid 2-disk $D^2\times S^1 \rightarrow S^1$). This idea of “shrinking” spaces is important because if a space $Y \subset X$ is a deformation retract of X, then $\pi_1(X) \simeq \pi_1(Y)$. Retraction can occur many ways: the space R^2 can be retracted to an open disk or a line or even a point. The 2-disk boundary ∂D^2 is not a retract of D^2 because the disk retains a point center. But once that center is removed,

we can even say that R^n-{0} →S^{n-1} using a retraction function r(x) = x/|x| to a unit sphere. Or, using the straight line progressive sliding idea from above, a point x on any inner circle of a disk, D^n- {0} can be expanded to its boundary using a mapping H(x,t) = (1-t)x + t x/|x|, a unit sphere when t = 1 (also called a deformation retract).
If a space has the homotopy type of a point, it is called contractible.

Homology Groups:

Homology reveals "holes" in any number of dimensions. Its objects are classes of k-dimensional "cycles" called k-cycles. Beginning with the question, Can loops be continuously deformed into each other. Its original goal was to study and classify holes in a manifold (like the 2-d plane E^2 or R^2). We learned in high school that the simplest concept of this was the **Euler characteristic,** chi = **χ** = V – E + F (adding number of vertices – edges + faces of a polyhedron). So, any spherical polyhedron has characteristic χ = 2 [e.g., four faced tetrahedron: χ = 4-6+4 = 2]. A torus has χ = 0, and a double torus (~thick figure "**8"**) has χ = -2. So number of holes =g= -χ/2 +1 (and this is also called the "genus," g). The homology group H_k(X) describes the number of k-dimensional holes in space X. The word "**hole**" is slightly unclear; but if you can put a string through it then it has a 1-d hole, and if you can put water in it then it has a 2-d hole. The sphere S^2 has a 2-d hole but no 1-d holes; so $H_2(S^2)$ = Z [but $H_1(S^2)$ = $\pi_1(S^2)$= 0]. A torus has two 1d holes (one for each S^1) and a 2-d hole that is the torus itself.. So $H_2(T^2)$ = Z and $H_1(T^2)$ = ZxZ (or $Z^2 = Z \oplus Z$).

This is also accounted for by something called the "Poincare polynomial" which for a circle is just (1+1x) where the x coefficient stands for 1-hole. The torus is $T^2 = S^1xS^1$ for polynomial $(1+x)^2 = 1 + 2x + 1x^2$ with 2 1-holes and 1 2-hole as coefficients (also called "Betti numbers). It would follow that a 3-torus T^3=$S^1xS^1xS^1$ would possess one 3-hole (even if we can't picture it). A 3-sphere S^3 would also possess a 3-hole [and $H_3(S^3)$) = Z, but $H_2(S^3) = H_1(S^3)$=0].

If two spaces X and Y are homeomorphic, H_r(X,Q) ~ H_r(Y,Q), and then χ(X)= χ(Y) and genus(X) = genus(Y).

Homology is also capable of describing geometries well beyond picturing such as complex projective space (CP^n).

There are three reasonable homology theories and go by the names: Singular, Simplicial, and DeRamm Cohomology.

The most intuitive approach to homology is to reconstruct an object using "simplices." A k-**simplex** is a generalization of having a single point in R^n be a 0-simplex, an open line interval is a 1-simplex, a triangle is a 2-simplex, and a tetrahedron is a 3-simplex. These are oriented building blocks put together to form complexes. The standard idea of a simplex is simply to form a convex figure to connect all the orthogonal unit vector tips with straight lines or flat planes. A 2-simplex is a triangle. For R^3: tips of i, j, k connected by lines and planes form a tetrahedron. Math can easily construct higher n-simplex in spite of our inability to visualize it. We say that H_n(X) is a simplicial homology group of a simplicial complex X using a simplicial chain complex C(X). Without this approach, computation can be quite difficult (but there are now software packages to do the task).

The space of a complex is a "polytope," and the complex is a "**triangulation**" of the complex. A "category" is a class of concepts and a class of morphisms. For every ordered pair (X,Y) of objects there is a set of morphisms Hom(X,Y). If the objects themselves are treated as categories, then the morphisms are called Functors. Homotopy is a category (e.g., there is a mapping between two curves having common endpoints). Topological invariants can utilize approximating geometric objects by polytopes (e.g., triangulating a torus to get Euler charactistics). (Reference Keesee).

Hilbert question: "Is every topological n-manifold triangulable?" Answer: No! (Kirby 1969), $\exists$ a 5-manifold (torus) which cannot be triangulated. And there is a 10-dim Manifold that does not admit a differentiable structure.

<u>Homology versus Homotopy:</u> It would appear that homology and homotopy capture the idea of "hole" equally well ($H_1 \sim \pi_1$). But, π_1 is not Abelian while H_1 is Abelian, so $H_1 \neq \pi_1$.

In higher dimensions, homology is a superior concept and they do different things. Note that $H_2(T^2) = Z$ but $\pi_2(T^2) = 0$. $\pi_3(S^2) = Z$, but $H_3(S^2) = 0$.

An annulus (a washer is a disk with a hole in the middle) and a half-twisted "Mobius" band are both homotopy equivalent to a circle, but they are not homeomorphic. When comparing the Mobius band and the cylinder, each can be smoothly shrunk by a homotopy to their equatorial circles. Yet one is an orientable surface and the other is not. Similarly a solid Klein bottle is homotopy equivalent to a solid torus ($D^2 \times S^1$). Note that a true Klein bottle cannot exist in 3-D so that our usual glass bottle picture is a Kludge. S^1 and an annulus are not homeomorphic, so homotopy is not sufficient to classify up to homeomorphism.

Why do the homology groups capture holes in a space better than the homotopy groups? A good interpretation of having an n-dimensional hole is a space X is that some image of the sphere S^n in this space given by a mapping $f: S^n \rightarrow X_f$ cannot shrink down to a point. The matter of "shrinking to a point" is best expressed by being ff homotopic to some constant map. Next, the homotopy groups π_n can be defined as the homotopy classes of base-point preserving maps from S^n to X. In this way it might be argued that the homotopy groups π_n should best capture the holes in X.

But this is not so. One has the most satisfying result that for $i \geq 1$ the homology $H_i(S^n)$ is nontrivial iff n=i. But the higher homotopy groups of spheres are very complicated. And a ball, D^3 [a 3-disk], is not homeomorphic to a point – but it is homotopically equivalent. The contraction all the way from the ball to the point doesn't violate anything about the homotopical equivalence.

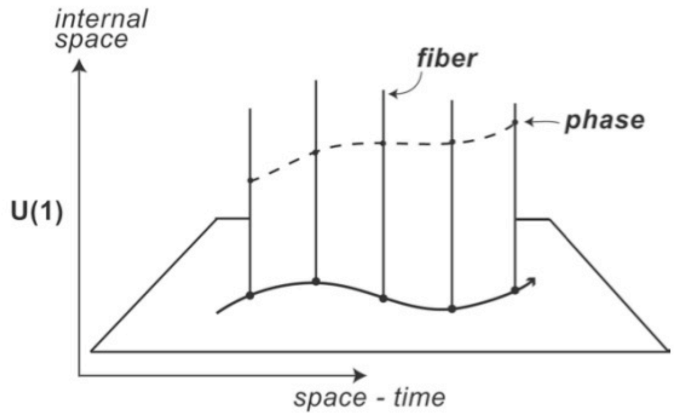


Figure 1: Phase Fiber Bundle with base space B = R^4 and U(1) phase fibers. The fiber segments have their endpoints identified so that they represent circles (standard figure).

Fibre-Bundles:

A "fiber-bundle" is a structure **(E, B, π, F, G)** consisting of a total topological space, E, a base space, B (usually connected and sometimes called "M"), a projection map from E to B (π: E→ B), a "fiber" F, and sometimes a structure group G guiding F [and when that applies, we use the name "principle bundle," **P**, for E, and structure (P,B,π, G)]. It is a generalization of the tangent bundle concept. In some modern physical cases, the fiber represents an "internal space" while the base B may be a "real" space. Figure 1 shows a quantum-mechanics example where an electron on a space-time trajectory has a phase that changes along its motion, and the phase lies in the 1-d unitary group U(1). A gauge group G can act on each fiber of the bundle separately.

The application of fiber-bundles is relatively recent. The first use of a fiber space may go back to Hassler Whitney (1935 using the name 'sphere-space'). More than mathematicians, physicists care strongly about "connections on principle bundles" and these may have first been applied by C. Ehresmann in 1950. Dennis Sciama may have been the first to picture gauge fields as connections on a fiber bundle (1958). The vector potential A is a U(1) connection for electromagnetism. A higher gauge theory was the Yang-Mills theory of 1954 which became very popular in the later context of the SU(3)×SU(3)×U(1) standard model.

An introduction to fibre-bundles commonly begins with the twisted **Möbius band** (Mö^2) where a base space is simply a circle, B = S^1 (usually pictured separately below the band, but it is also the centerline of the band about the loop). The band usually has some thickness (perpendicular to the edges) that we could call interval I =[0,1], and each width-line of thickness is a fiber, F. Each small neighborhood of the band, U ⊂ E can be mapped back onto the base circle, mapping π: E→ B. Each point of the base space, say angle θ, points to a fiber, π^{-1}: B→ F, θ → all y ∈ [0,1]. Locally, the Möbius band is a "product-space" B×F (like a cylinder), but the band has a twist that is only visible globally. The cylinder has a global product C ≃ S1×[a,b] and so is a trivial bundle with fiber F = I ⊂ R. But the Möbius strip, Mö^2, is the simplest example of a nontrivial bundle and requires two circles as a cover (continuous non-intersecting double loop). A principal bundle is trivial if and only if it allows a global section.

In physics applications, the base manifold B may be **space-time** itself, M^4 (as in Figure 1). For quantum mechanics and the de Broglie phase of a "particle," we picture a little circle attached to points $(x,y,z,t) \in B$. The fiber is $F=S^1$ with structure group G=U(1). For G=SU(2), we picture a little sphere attached to every point $p \in B$ (not quite appropriate). These are also the pictures that Brian Green uses to represent some of the extra compact dimensions associated with string-theory.

<u>Wikipedia:</u> "An elegant and intuitive way to formulate Maxwell's equations is to use complex line bundles or principal bundles with **fibre U(1).** The connection ∇ on the line bundle has a curvature $F = \nabla^2$ which is a two-form that automatically satisfies $\mathbf{d}F = 0$ and can be interpreted as a field-strength. If the line bundle is trivial with flat reference connection d we can write $\nabla = \mathbf{d} + A$ and $F = \mathbf{d}A$ with A the 1-form composed of the electric potential and the magnetic vector potential.

For the Lie group SU(2), there is tan**gent space su(2)** forming a tangent bundle.

<u>**"Section:"**</u> *[from the concept of "cross-section" but not necessarily planar].*

A "section" of a fiber bundle is a continuous map, s, from base manifold B to a particular set of chosen points of the fibers. That is, for fiber bundle $\pi: E \rightarrow B$, compose $\pi(s(x)) = x$ = identity for all $x \in B$. As a simple function example, temperature over space and time: T = s(x,t) where s <u>"lifts"</u> points on the base space+time manifold to values of their temperature and is a section of the trivial line bundle B×R. We are selecting just one point for each fiber F = R. If we think of each fiber as possessing a zero value, then the zero section selects all zero points over all fibers. The phase line in Figure 1 is a section.

A more interesting but still simple vector bundle example might be the values of the **vector potential, A**(r) **,** versus radius, r, for the case of an ideal (infinite) <u>solenoid</u>. The base manifold is just a plane normal to and cutting through the solenoid and includes all points $(r, \phi) \in B$. Inside the solenoid, the magnetic field is <u>constant</u>, and $A = A_\phi = \mu_o nir/2$ ramping up from zero to peak on the coil located at r = R. The magnetic field outside the coil is zero but with a circular A field falling off as $A_{out} = \mu_o niR^2/2r$. The joint plot of $A = A_{in} \cup A_{out}$ versus r is the section profile graph and depends on the current flow i. For all circles in B about a center r=0, all A's are tangent vectors to the circles. This fiber bundle has a "sheaf" of sections each parameterized by current, i.

In wave mechanics, the base space M is configuration space (space of n distinct points in M), fibers are Hilbert space, H, and the wave-function psi is a section [25]. A gauge transformation is a change of section.

<u>Connection and Curvature</u>:

A "connection" defines <u>parallel transport</u> on a bundle and is equivalent to a "covariant derivative" for vector bundles. It is not unique, there are different connections for different purposes. The idea of two vectors being parallel depends on the specific path joining their two points. In general relativity, "connection coefficients $\mathbf{\Gamma}^j_{mn}$ serve as

turning coefficients to tell how fast to turn the components of a vector in order to keep that vector constant (against the turning influence of the vectors)." [MTW p 212]. This use of Christoffel symbols applies to the "Levi-Civita" connection on Riemannian manifolds. The difference between a partial derivative (or coordinate derivative) and a covariant derivative is the connection correction. The notion of curvature says that if we attempt a parallel transport about a parallelpiped, the beginning and end point won't match up. These ideas can be abstracted beyond metric spaces to include even electromagnetism. The electromagnetic Schrodinger equation, for example, uses a momentum operator $p_\alpha \sim [-i\hbar\,\partial_\alpha - eA_\alpha]$ so that $\nabla_\alpha = \partial_\alpha - (ie/\hbar)A_\alpha$, where the last term is the connection correction for the covariant derivative for the wave-function bundle. It is a 1-form ω_α [Frankel, p.442]. And then the curvature of the connection is the EM 2-form: $\theta = \mathbf{d}\omega + \omega\wedge\omega = \mathbf{d}\omega = -(ie/\hbar)\,\mathbf{d}A = (-ie/\hbar)\mathbf{F} = -(ie/\hbar)[E\wedge dt + \beta]$.
(because U(1) is Abelian so that $A\wedge A = 0$). Another way to state this is: "Out of A we can construct only two possible 2-forms: dA and $A\wedge A$." So in general F must be a linear combination of the two." [Zee]

We say a manifold M is curved if its "tangent spaces $T_p(M)$, $T_{p'}(M)$ at two neighboring points p and p' change as one moves from p to p' [Nash, p174]. A connection is essentially a structure which endows one with the ability to compare two such tangent spaces at a pair of infinitesimally separated points. The connection is given by defining what is called parallel transport…" A connection may be expressed by covariant differentiation. "Being unlike partial differentiation, this will not in general be commutative." And a measure of the non-commutativity of covariant differentiation is the curvature. "Every connection can be shown to arise from a certain 1-form ω belonging to T*P."

At any point q, the tangent space T_qP to a principal bundle P may be decomposed into two disjoint subspaces called the "**vertical**" subspace V_qP parallel to a fiber and a "**horizontal**" subspace H_qP transverse to a fiber so that $T_qP = V_qP \oplus H_qP$ [24]. The horizontal space is also key to what is called the "Ehresmann connection." The group G acts to push points forward along fibers: $\Phi(g,q) = \Phi_g(q) = q\cdot g$, where $g \in G$. This action may also be used to push the horizontal subspace along the fiber away from the base space: $T_qP \rightarrow T_{\Phi(q)}P$. A connection defines flat horizontal subspaces near q with the above properties and isomorphic to T_qM (the base space). A connection defines a "horizontalized" version of the exterior derivative [25].

When moving up in complexity from the usual U(1) electromagnetism gauge group to an G=SU(2) Yang-Mills theory, the connection and curvature both become more complex and require more indices: gauge potential $A^\alpha{}_\mu(x)$ and gauge field tensor $F^\alpha{}_{\mu\nu}$. The characteristic class tells how far a bundle is from being trivial. And in this case, curvature F over Pauli matrices, $F = F^\alpha \sigma^\alpha/2i$ has non-vanishing wedge product $F^\alpha \wedge F^\alpha$. This in turn gives non-vanishing 2nd Chern class, c_2.

A cross section that is "constant" is also called **horizontal**. But with a connection, it may be path dependent.

Integration: Stokes' Theorem:

In standard Vector Analysis, Stokes' Theorem (1850) says that if a surface S is open, 2-sided, and bounded by a simple closed curve C, then if field A is differentiable:

$\oint_c A\cdot dr = \int_s (\nabla \times A)\cdot \mathbf{n}dS$. We also have the Gauss Divergence Theorem:
$\int_v \nabla\cdot EdV = \int_s E\cdot \mathbf{n}dS$, for a volume V bounded by a closed surface S (1762,1813).

These can be simplified and generalized to n-dimensions using the language of forms: "Let $V^p \subset M^n$ be a compact oriented submanifold with boundary ∂V in M^n. Let ω^{p-1} be a continuously differentiable (p-1)-form. Then
$\int_V \mathbf{d}\omega^{p-1} = \int_{\partial V} \omega^{p-1}$. E.g., for the magnetic field 2-form β, $\int_U \mathbf{d}\beta = \int_{\partial U} \beta = 0$. Since this is true for arbitrarily small neighborhoods U, it must be that $\mathbf{d}\beta = 0$ (translation, $\nabla\cdot B = 0$, no poles). Or Faraday's Law: $\oint_{\partial V} E^1 = -\int \partial\beta/\partial t$ gives $\mathbf{d}E^1 = -\partial\beta/\partial t$, Translates to: curl $E = -\partial B/\partial t$.

Other preliminary concepts:
Trace, the trace of a square matrix A is $Tr(A) = \Sigma a_{ii}$, the sum of all diagonal elements. The trace of the product of two square matrices: Tr(AB) = Tr(BA). This bears some similarity to the dot product of vectors. So, a generalization of vector operations to matrices often involves a trace of matrix products.
And trace is invariant under similarity transformations; that is, similar matrices have the same trace. This invariance under basis change is valued in Math and Physics. Trace is a linear functional so that tr(cA + B) = c tr A + tr B. Also Tr(ABC) = Tr(CAB) = Tr(BCA) under cyclic permutation. $Tr(A\wedge B) = Tr(B\wedge A)$ only if degA x degB is even (else -).

Gauss Bonnet Theory:

"In 1944, fiber bundle theory became important in topology with Chern's generalization of the Gauss-Bonnet theorem to four dimensions." The standard Gauss-Bonnet theorem says that $\int_M KdA = 2\pi\chi(M)$ where χ is the Euler-Poincare characteristic and $\int KdA + \Sigma\alpha_i = 2\pi$ where α is an "exterior" angle or "turning angle." For example, a flat equilateral triangle has interior angle $\theta = \pi/3$. The angle from a flat edge then turns $180^o - 60^o = 2\pi/3$. So, $2\pi\,\chi(M) = 2\pi - \Sigma\alpha_i = 2\pi - 3(2\pi/3) = 0$, i.e., $\chi(\Delta) = 0$.
Also, $\chi(\Delta) = v - e + f = 3-3+0 = 0$ (check). On S2, great circle triangle from N-pole to equator around 90^o and back to north has equal 90^o corner angles and surface area of $1/8^{th}$ of $4\pi R^2$. Gaussian curvature is $K = 1/R^2$, so:
$\int KdA + \Sigma\,\alpha_i = (1/R^2)(4\pi R^2/8) + 3(\pi/2) = \pi/2 + 3\pi/2 = 2\pi$.
For a whole sphere, $\chi(S^2) = 2$ and for a torus, $\chi(T^2) = 0$. Also for a sphere, polyhedral like a tetrahedron cube have v – e + f = 4 – 6 + 4 = 2, or 8-12+6 = 2, i.e., $\chi(S^2)=2$. And K Area(sphere) = $4\pi = 2\pi\,\chi(S^2)$, $\chi = 2$.
The Euler-Poincare characteristic is a prototypical integer index for characterizing spaces.

A simple example of the Gauss-Bonnet theorem is the Berry phase of the precession of a **Foucault** pendulum in a lab at a given latitude as the earth spins a full daily rotation. A full precession occurs at the north pole, but there is no precession at the equator. The easiest calculation for this is via the Gauss-Bonnet theorem saying that phase shift is the same as the enclosed solid angle of a spherical cap bounded by the given latitude, θ. The easy integral of $\alpha \sim \Omega = \int_o^\theta 2\pi R\sin\theta d\theta$ gives an angle $\alpha = 2\pi(1-\cos\theta)$ [e.g., at polar angle (from N pole to θ) $\theta_{polar} = 90 - 40^o$ for Boulder, Colorado, $\alpha = 128^o$ or 2.24 radians/day]. The latitude circle path, C, is not a great circle geodesic, so turning angles result from parallel transport of some initial vector direction from a geodesic propensity. Another approach is look at the set of all north pointing tangent lines along a latitude path C. This is obviously a cone with an apex A lying directly above the north pole and having side length a = Rtanθ. The circumference C = 2πR sinθ = 2πρ where ρ is the circle radius from the spin axis. If a cone is cut along one side and then stretched out flat, there will be a missing wedge with angle α . If an initial vector V_o points along the latitude, it will rotate in going about C on the tangent cone. On the flat projection, V(ϕ) will always be parallel to V_o. But after precessing along C, it will end up with a twist of angle α. The wedge is the angular difference between a full flat circle of radius a and the circumference C. This ΔC = 2π (a-ρ) = 2π a(1-ρ/a) = 2π a(1-cosθ), or ΔC/a = 2π(1-cosθ) – again.

<u>CHERN Classes</u>:

Shiing-Shen Chern was born in China in 1911. In the summer of 1934, Chern graduated from Tsinghua with a master's degree, the first ever master's degree in mathematics issued in China { pinyin: Chén Xǐngshēn}. He moved to the University of California, **Berkeley**, in 1960, where he worked and stayed until his retirement in 1979. In 1961, Chern became a naturalized citizen of the United States; and, in the same year, he was elected member of the United States National Academy of Sciences.

<u>Names</u>: There are Chern Classes [c_i(V)], Chern numbers (e.g., c_1 c_2), form [c_k(V)], Character ch(V), polynomial c_t(E), and roots. Chern classes were introduced by Chern in 1946.

Topological characteristic classes are often studied in the language of forms, and there are four major classes: Chern, Pontrjagin, Euler, and Stiefel-Whitney [Kaku book on String Theory]. Popular examples include the group SU(2), and Chern forms are generally taken for a U(n) Bundle. Characteristic is related to **eigenvalue polynomials**, det(λ I –A)=0 which is invariant in the sense that "similar matrices have the same characteristic polynomial". For det(I+A) simply let λ = -1 in det(λI – A) characteristic polynomial, then replace A by a matrix of curvature 2-forms θ i/2π to get Chern form coefficients.

The differential forms are polynomials in F where F is a curvature 2-form for the bundle P and connection A determines the curvature. The polynomials are invariants of the Lie algebra g of G. The polynomials P_i(F) is independent of the connection A used to compute F.
Electromagnetism is a connection on a U(1) bundle. Without a monopole, U1 is trivial; and with a monopole, U(1) is nontrivial. Dirac monopole quantization is a classification of a U(1) bundle according to a first Chern class.
"a principle SU(2)-bundle over a 4-manifold X has a second Chern class $c_1 \in H^4(X,Z)$. [superscript means cohomogy group]. Every principle SU(2) bundle over M = R^4 or M = R^4-{0} is trivial. In general, for fiber bundle E, $c_k(E) \in H^{2k}(M,Z)$. [double check?].

One example of Chern number application is the commonly applied "Berry phase." Berry phase (1984) or <u>"geometrical phase"</u> is a difference in phase that is acquired over the course of an adiabatic cycle about a closed path, C. It can be found by integrating the "Berry connection" around the loop C or by integrating the "Berry curvature" over a surface enclosed by C:

$\oint A_\mu (x)dx^\mu = \int F_{\mu\nu}(x)\, dx^\mu \wedge dx^\nu = 2\pi$ (integer Chern number) for the first Chern class. For the special case of the important experimentally verified "Aharonov-Bohm" phase shift result for electrons moving about a solenoid, the A really is the vector potential and the F really is the electromagnetic tensor. The "nonintegrable" geometrical phase can be described roughly as "global change without local change" and has become so popular that it has been called "the phase that launched 1000 scripts." The origin of the Berry's phase is in nonflatness of a parallel transport which appears in the corresponding phase factors and may be described by holonomy in fiber bundle theory.

CHERN-SIMMONS, CS, and "ada" :

The Chern–Simons theory, named after Shiing-Shen Chern and James Harris Simons, is a 3-dimensional topological quantum field theory of "Schwarz" type further developed by Edward Witten. It is so named because its action is proportional to the integral of the Chern–Simons 3-form. "CS forms were originally introduced in physics in the discussion of chiral anomalies. Chern-Simons theory is called a topological gauge theory because it is a gauge theory that does not require a metric.

In the case of topological quantum field theory and integral or fractional Hall effects, Chern-Simmons theory stresses a single famous term "ada." (Zee).
With indices this is: AdA = ada = $\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a$ and applies to "2+1" spaces where the "1" is time with an a_o term. The ada term also appears in the CS 3-form. [ϵ is the Levi-Civita anti-symmetric tensor].

A recent note in "Quora" said, "Dirac's intuition predated Chern's formal definition of the Chern class by roughly twenty years. In fact, one of my advisors told me that at a conference in Stony Brook in the late 70s, Chern spoke of how Dirac's discoveries inspired him to consider the objects that would later become known as Chern-Simons form." There are Chern-Simmons forms, actions and invariants and they can deal with fractional statistics or anyons [8] .

In condensed matter physics, Chern–Simons theory describes the topological order in fractional quantum Hall (FQH) effect states. In mathematics, it has been used to calculate knot invariants and three-manifold invariants such as the Jones polynomial.
A recent paper [10] said, "Loosely speaking, three dimensional Chern-Simons theory is the theory of an integral called Chern-Simons action, of some "characteristic" differential form defined over the spaces of connections on 3-manifolds with values in a fixed Lie algebra."

The 2-form curvature is = $\theta = \mathbf{d}\omega + [\omega, \omega]/2 = \mathbf{d}\omega + \omega \wedge \omega$ (older notation is $\theta^a{}_b = ½ R^a{}_{bjk}\, dx^j \wedge dx^k$). For EM, the connection $\omega = -iqA/\hbar$.
"The topological significance of $\mathbf{tr}\ \boldsymbol{\theta} \wedge \boldsymbol{\theta}$, generalizing Poincare's theorem for closed surfaces was discovered by Chern, and these types of integrands are called Chern forms and symbolized by c_r.
$c_2(E)\ (1/8\pi^2)\ \mathrm{tr}(\theta \wedge \theta)$ which is the 4-form appearing in the winding number of an SU(2) instanton, and E or P is a principle bundle (over M with group G).

4-form from d [CS 3 form]: tr(θ∧θ) = **d** tr [ω∧dω + (2/3)ω∧ω∧ω) for any vector bundle and is proportional to the Chern-Simons number N_{CS}. [Frankel, p 586]. Or, depending on how A is defined, **d** tr(AdA+2A^3/3). [again with the ada term].
[Check: to show this, let θ =dt∧ω + tdω + t2ω ∧ω with parameter t.
Then expand θ∧θ into a collection of terms. Then take the exterior derivative d of the claimed result and get same collection of terms. The denominator /3 gets removed by three terms for d(ω∧ω∧ω). We are using concepts like d(dω) = 0 and
Tr(ω∧ω∧ω∧ω) = 0 – ∧ 3 times may be non-zero, but 4 times gives zero].

Baryon and lepton numbers are not exactly conserved quantities in the Standard Model, because of the axial "anomaly" (violation of the classical conservation of the axial current) that connects them to the Chern-Simons number of the weak gauge field. Vacua in the electroweak theory are labeled by an integer-valued Chern-Simons number, N_{cs} . Particularly, Chern–Simons theory is specified by a choice of simple Lie group G known as the gauge group of the theory and also a number referred to as the level of the theory, which is a constant that multiplies the action. The action is gauge dependent, however the partition function of the quantum theory is well-defined when the level is an integer and the gauge field strength vanishes on all boundaries of the 3-dimensional spacetime.
The Atiyah-Singer Index theorem is also expressed using this Chern character.

The general EM Lagrangian is L = ¼ (F∧*F + ΘF∧ F) where the theta vacuum slight modification is often deleted.

Monopoles, instantons, sphalerons, anomalies:

Monopoles: First, as simple example, consider the winding number of a function map as a topological invariant or "charge." For points on a circle, a smooth function full rotation, ϕ(2π), could be the same as ϕ(0) on another circle, as it might for the strict definition of the word "function." But it could also be 2πn where n is a winding number integer. This is covered by the homotopy $\pi_1(S^1) = Z$. And, for the number of distinct ways that points of a sphere can be smoothly mapped onto points of another sphere (19), we also know that the topologically distinct ways this mapping can be done is labeled by $\pi_2(S^2) = Z$. And for a 3-sphere, it is also true that $\pi_3(S^3) = Z$.
For a sphere surrounding a magnetic monopole (Dirac proposal, 1931), a mapping of S^2 onto S^2 from 2-d space to isotopic space also obeys $\pi_2(S^2) = Z$. The Dirac monopole is a topological defect in a compact but not simply connected U(1) gauge theory, and its magnetic flux is the first Chern number of the principle bundle, c_1.

There are many derivations of Dirac quantization for magnetic monopoles (e.g., Kaku p. 541). Perhaps one of the clearest was given by Wu and Yang (see **Zee** [8] p. 220): We assume the existence of a magnetic monopole with magnetic pole charge g (although no experimental discovery is yet claimed). This has a radial field pointing away from the charge: $B = g/4\pi r^2$ and hence a magnetic flux through a surrounding sphere $\oint B\cdot da = \int B(r \sin\theta d\phi\, r d\theta) = 4\pi r^2 B = g$.
This is similar to the Gauss and divergence theorem for electric field, E.
$\int \nabla\cdot E\, d(vol) = \oint E\cdot da = (q/4\pi\epsilon_o r^2)(4\pi r^2) = q/\epsilon_o$.

The term under the integral sign, $Br^2\sin\theta d\theta d\phi = Br^2 d\cos\theta d\phi$ is also a magnetic 2-form that could be labeled as $F = (g/4\pi)d\cos\theta d\phi$, and again $\oint F = g$ [of course F already includes the increment of area].
If this 2-form is the exterior derivative of a 1-form gauge potential, $F = \mathbf{d}A$, then $A = (g/4\pi)\cos\theta d\phi$ [and remember that $dd\phi = 0$].
But, $d\phi$ is <u>not defined</u> at the poles N and S. So we have to have two patches with each avoiding the upper or lower z-axis on the space $R^3 - \{0\}$. Then we can create A potentials $A_N = (g/4\pi)(\cos\theta - 1)d\phi$ [still good for $\theta = 0$] and $A_S = (g/4\pi)(\cos\theta+1)d\phi$ [still good for $\theta = \pi$]. And each of these still satisfies $F = dA$.

Now apply some ideas from gauge invariance for U(1) and quantum mechanics: Let $\Lambda(x)$ [or sometimes $q\chi(x)/\hbar$, but with $\hbar \equiv 1$] for $\psi(x) \rightarrow \psi(x)\exp(i\Lambda(x))$ with the accompanying "compensating field" $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \Lambda(x)/q$.
This added term $\partial_\mu \Lambda(x)/q = (iq)^{-1}\exp(-i\Lambda)\,\partial_\mu \exp(i\Lambda) = (iq)^{-1}e^{-i\Lambda}\,\mathbf{d}e^{i\Lambda}$ using forms. So, for a gauge transformation , we now need to equate $2(g/4\pi)d\phi = (1/iq)e^{-i\Lambda}\,\mathbf{d}e^{i\Lambda}$ so that $e^{i\Lambda} = \exp(i2(qg/4\pi))\phi$. The scalar function $\Lambda(0) = \Lambda(2\pi)$, so $i2(qg/4\pi)(2\pi)=1$ or $e^{iqg}=1$.
For the electron, q = e, so <u>**$g = 2\pi\hbar n/e$**</u> or $e = 2\pi\hbar n/g$ for usual SI units [Dirac quantization, and showing $\hbar$ put back into the equation]. [For Gauss-cgs units, it wold be $e = \hbar cn/2g$].
A lesion is "that F is locally but not globally exact – where F = dA being exact would require g = 0. A small charge for e implies a very large charge for q.

<u>Instanton</u>: A Lagrangian for a quantum tunneling solution may plot as a very short <u>instantaneous</u> blip versus time and is hence called an instanton (19). The instanton solution of the Yang-Mills equations was discovered by Polyakov in 1975 and later attributed to a tunneling event between degenerate classical vacua. The term instanton was coined by 't Hooft [22]. Instantons can probe the nonperturbative realm of gauge theories such as Yang-Mills (19). And QCD Instantons play a role in chiral symmetry breaking [22]. They are localized (e.g., 1/3 fm) regions of space-time with very strong gluonic fields.

Evaluating the integral of a 2nd Chern type form may have fields vanishing slowly enough at infinity to yield non-zero values and hence winding numbers for finite action. Evaluated over a hypersphere boundary gives a degree of mapping from $S^3 \rightarrow S^3$. The Yang-Mills instantons then give topologically distinct vacua each labeled by an integer n (19). The winding number of the instanton is $-\int c2$ over R^4 for SU(2) bundles (Frankel). Instantons are topologically nontrivial solutions of Yang–Mills equations that absolutely minimize the energy functional within their topological type.

Sphaleron: This word is taken from the Greek language and means "ready to fall", which resembles the fact that sitting on top of a rounded minimal energy curve is "slippery." The application is electroweak symmetry breaking and baryon non-conservation using the gauge group SU(2). This group can lead to the existence of topological effects some of which are classified by an integer topological winding number (N_{CS} Chern-Simmons number).

Change in <u>Chern-Simmons number,</u> transition creates 9 left-handed quarks (3 colors x 3 generations) and 3 left handed leptons (one per generation). If the system is able to perform a transition from the vacuum $G_{vac}(n)$ to the closest one $G_{vac}(n\pm1)$, the Chern-Simons number is changed by unity and $\Delta B = \Delta L = nf$. That is, changes in Chern-Simons number result in changes in baryon number which are integral multiples of the number of families nf (with nf = 3 in the real world). Tunneling and vacuum transitions can yield baryon number non-conservation, q+q→ 7qbar + 3 ℓ bar. ?? Gauge transformations U(x) which connects two degenerate vacua of the gauge theory may change the Chern-Simons number by an integer n, the winding number. The periodic sphaleron humps leading to B+L violations may resemble a sine wave with peaks at an energy of 9 TeV! Probing these processes may be done using ultra-high energy neutrino events as seen by cubic-kilometer neutrino telescopes. In particular, a study of IceCube sensitivity is seen to be similar to that of the first 13 TeV LHC data. Chern–Simons number N_{CS}, is an integer for vacuum configurations where topologically different bosonic ground states are separated by an energy barrier. The sphaleron configuration are on top, with half-integer N_{CS} and an energy of about 9 TeV.

"The baryon number is violated in the Standard Model by non-perturbative sphaleron transitions. At temperatures above the electroweak scale, the rate of the sphaleron transitions is unsuppressed and has been accurately measured using effective theories on the lattice. At temperatures substantially below the electroweak scale, the Higgs field expectation value is large and the sphaleron rate is strongly suppressed. Here analytical estimates are sufficient. The sphaleron rate, however, has not been calculated in the intermediate temperature range with physical Standard Model parameters." One work uses "an effective electroweak theory on the lattice with multicanonical and real-time simulation methods to calculate the sphaleron rate through the electroweak crossover at Higgs masses of 115 GeV and 160 GeV. (ref date?)

Transitions between vacua are possible by surmounting the potential barrier through sphaleron transitions. The sphaleron rate is strongly suppressed at low temperatures, where the potential barrier is high. At temperatures above the EWPT, though, transitions among vacua are made possible because there is no longer any potential barrier [10,11].

[WIKipedia] "Baryogenesis within the Standard Model requires the electroweak symmetry breaking be a first-order phase transition, since otherwise sphalerons wipe off any baryon asymmetry that happened up to the phase transition, while later the amount of baryon non-conserving interactions is negligible.
If a thermodynamic quantity changes discontinuously (for example as a function of temperature) then we say that a first order phase transition has occurred. EWBG requires first order <u>or topological defects</u>. The minimal standard model has neither enough CP-violation nor a sufficiently strong phase transition to allow electroweak baryogenesis to take place." That is, EWB doesn't work in the Standard Model, and EWB in the MSSM is almost ruled out The 125 GeV Higgs boson will be too heavy to give rise to a first order EWPT. A strongly first-order EWPT requires new Higgs interactions with particles beyond the SM. If physics up to the TeV scale is completely described by the SM, it is well known that the electroweak phase transition (EWPT) is second-order (continuous). The baryon number violation due to sphaleron transitions, is based on the non-Abelian SU(2) part alone. One can easily decouple the U(1) sector by setting the Weinberg angle to zero, $\theta_W = 0$. This disentangles the SU(2) and the U(1) parts completely.

Anomaly: Wikipedia says, "In quantum physics an anomaly or quantum anomaly is the failure of a symmetry of a theory's classical action to be a symmetry of any regularization of the full quantum theory." A standard example of an anomaly is the (anomalous) decay time of the neutral pion, π^o. Charged pi-mesons, $\pi^\pm$ have a lifetime of 26 nanoseconds and decay via the weak force. The pi-zero with mass 135 MeV decays electromagnetically and only lives for $8.4x10^{-17}$ seconds (about a thousand times shorter than predicted by older theory). The cause of this is a Feynman diagram with a closed triangular massive quark fermion flow between an input π^o vertex and 2 photon output vertices. Roman Jackiw used an older Steinburger idea of a proton triangular current but now using QCD. He found that his anomaly used the Atiyah-Singer index theorem of 1963 which in turn used a Chern character (proportional to trace(FF)).
The Chern class of a gauge field configuration is called a topological charge which for the pion anomaly is the difference of the number of right handed n^+ and left-handed zero modes of a Dirac operator.

References

1. Barrett O'Neill, Elementary Differential Geometry, Academic Press, N.Y., 1966.
2. Theodore **Frankel**, The Geometry of Physics, An Introduction, 2nd Edition, Cambridge, 1997, 2004.
3. Charles **Nash** and Siddhartha Sen, Topology and Geometry for Physicists, Academic Press, 1983. {An elaboration is: C. Nash, "Topology and Physics – a historical essay," 1997, https://arxiv.org/pdf/hep-th/9709135v4.pdf, }.
4. John W. Keesee, An Introduction to Algebraic Topology, 1970
5. π vs H: https://www.physicsforums.com/threads/geometric-difference-between-a-homotopy-equivalance-and-a-homeomorphism.621977/
6. "A Brief Introduction to Characteristic Classes from the Differentiable Viewpoint." http://www.lepp.cornell.edu/~yz98/notes/A%20brief%20introduction%20to%20characteristic%20classes%20from%20the%20differentiable%20viewpoint.pdf
7. Thesis details, Sphaleron EW 260 pages. http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/SchaldachJoerg/diss.pdf
8. Tony **Zee**, Quantum Field Theory in a Nutshell, Princeton, 2003.
9. John Ellis, "Search for Sphalerons: IceCube vs. LHC, arXiv:1603.06573 [hep-ph] 21 March 2016.
10. Stephane Baseilhac, Chern Simons Theory in Dimension Three, 59 pgs, CS.pdf, http://www.math.univ-montp2.fr/~baseilhac/CS.pdf
11. Zanelli, Chern-Simmons theory in dimension three. http://pos.sissa.it/archive/conferences/124/004/ICFI%202010_004.pdf
12. Michel Laurent, Fibre bundles, 9/13/2007.
13. Ralph L. Cohen Dept. of Mathematics, The Topology of Fiber Bundles Lecture Notes, 178 pages. http://math.stanford.edu/~ralph/fiber.pdf
14. Morihiro Honda, Fiber Bundle Structure of Gauge Field and Reduction by the Higgs Mechanism, Prog. Theo. Phys. 63, 4, April 1980, p 1429, http://ptp.oxfordjournals.org/content/63/4/1429.full.pdf
15. MTW: Misner, Thorne, Wheeler, **GRAVITATION**,
16. Harley Flanders, Differential Form, with applications to the physical sciences, Dover publications, 1963, 1989.

17. Richard L. Bishop and Samuel I. Goldberg, Tensor Analysis on Manifolds, Dover Publications, 1968, 1980.
18. Roger Penrose, The Road to Reality, a complete guide to the laws of the universe, Knopf, 2004, 1098 pages. Chap. 15, Fiber Bundles and gauge connections.
19. Michio Kaku, Quantum Field Theory a modern introduction, Oxford, 1993, Chapter 16.
20. M. Daniel and C.M. Viallet, "The geometrical setting of gauge theories of the Yang-Mills type, Reviews of Modern Physics, RMP Vol. 52, No. 1, January 1980, p 175.
21. William L. Burke, Applied differential geometry, Cambridge, 1985.
22. T. Schafer, "Instantons in QCD", ArXiv 9610451, 140 pages
23. Karl F. Warnick, Peter Russer, "Two, Three and Four-Dimensional Electromagnetics Using Differential Forms," 2006, see: http://journals.tubitak.gov.tr/elektrik/issues/elk-06-14-1/elk-14-1-14-0509-13.pdf Or: "Differential Forms and Electromagnetic Field Theory," 2014 Progress in E. Res., 148, p83-112, http://www.jpier.org/PIER/pier148/09.14063009.pdf .
24. Rupert Way, "Introduction to connections on principle fibre bundles," 3/24/10, Surrey UK. http://personal.maths.surrey.ac.uk/st/T.Bridges/GEOMETRIC-PHASE/Connections_intro.pdf Hp and Vp subspaces.
25. Advanced Electromagnetism: Foundations, Theory and Applications edited by Terence W Barrett, Dale M Grimes.
26. Nima Arkani-Hamed and Jaroslav Trnka, The Amplituhedron, 2013, http://arxiv.org/pdf/1312.2007.pdf
27. Manuel Asorey, "Space, matter and topology," for Nature Physics 12, 616-618 (2016) or http://arxiv.org/pdf/1607.00666.pdf

28. David J. Jackson, "Unification in One Dimension," arXiv:1606.09568, 488 pages +. 16 Feb 2016.
29. Charles Kittel, Introduction to Solid State Physics, Third Edition, 1966, John Wiley and Sons. (p 253+)
30. H.P. Myers, Introductory Solid State Physics, 1990-1994, Taylor & Francis. (pages 120+ and 194+).

Appendix:

Additional Notes:

Nima Arkani-Hamed (*Dr. "Space-time is doomed*") introduced a new physics concept called the **Amplituhedron** [$A_{n,k,L}$] as something similar to a "volume" enabling direct computation of scattering amplitudes [26]. This is new abstract mathematics using twistor space but is also a strong simplification of calculations more typically done using Feynman diagrams. It uses an algebraic geometry "positive Grassmannian" similar to a convex polytope in projective space. Examples of Grassmannian Gr(r,V) for vector space of dimension r include: Gr(1,V) as a space of lines through the origin and Gr(2,3) as the space of all planes through the origin. He also introduced a new type of form: **dlog-forms** as rational form $\Omega = (d\alpha_1/\alpha_1) \wedge ()\ldots = d\log(\alpha_1) \wedge \ldots d\log(\alpha_{4L})$ for space of loop momenta and having only "logarithmic singularities." These apply to one-loop amplitudes like scalar bubble, triangle, and box integrals free of any poles at infinity.

For EM identify A with a connection on U(1) bundle: the vertical automorphisms of the bundle will reproduce the gauge transformation of electromagnetism. And F = dA turns out to be the curvature of the connection.
Any non-gravitational theory can be formulated on a fiber bundle associated with the principal bundle determined by the metric and connection: The Φ (x) break up into two subclasses: The fields of massive objects (such as charged bodies) are represented by geometric quantities living on the vertical fibers; and the gauge fields transmitting the forces between these objects (such as the electromagnetic field) are represented by verical connections along the fibers; these connections are only fixed up to some group of gauge transformations.

The structure group that defines the fibers of both bundles is U(1)-- the set of rotations in the complex plane, parametrized by the angle of rotation. In any principal bundle, the elements of the fiber are just the members of the structure group itself, whereas the fiber of the associated vector bundle consists of a vector representation of that group. In the case of electromagnetism, a vector in a fiber of the associated vector bundle is just a complex number. This is the value of the particles' position-representation wave-function at the space- time point lying "below" that fiber. A change of phase of the wave-function at that point corresponds to a rotation in the fiber "above" that point.

The connection on the principal fiber bundle representing electromagnetism is a geometric object: specifically, it is given by a Lie-algebra-valued one-form field on the bundle. It is defined independently of any choice of coordinate charts or section for the bundle. The pull-back corresponding to each local section on this bundle uniquely defines a one-form (or co- vector) field on a corresponding open set of the (space-time) base manifold M. The usual quantity A_μ is just a coordinate representation of this one-form field, but for a trivial constant factor. The usual quantity $F_{\mu\nu}$ is similarly related to a coordinate representation of the two-form field on M given by the pull-back of the bundle curvature.
On the Reality of Gauge Potentials, Richard Healey Philosophy Department, University of Arizona, http://philsci-archive.pitt.edu/328/1/RLGAUG%2Bfiguresfinal.pdf
Printed p 12-16.

Quora: As far as physicists are concerned, non-euclidean geometry and topology are not "advanced mathematics", they are ancient history, only slightly more recent than euclidean geometry and calculus. They were advanced mathematics in 1916. In 2016 they are (together with slightly more advanced fields such as differential geometry) on the list of bare essentials without which one can't even begin to do theoretical physics.

Notes: Sphaleron t'Hooft tunneling through 10 TeV barrier. EW symmetry is restored at 100 GeV.
[9, John Ellis, "Search for Sphalerons: IceCube vs. LHC, arXiv:1603.06573 [hep-ph] 21 March 2016]
The Chern-Simons number as an order parameter: classical sphaleron transitions for SU(2)-Higgs field theories for m_H ~ 120 GeV.
FROM FEB: The Θ Vacuum
The ground state of a quantized non-Abelian Yang-Mills gauge theory is usually described by a real-valued parameter θ a fundamental new constant of nature.
The structure of this vacuum state is often said to arise from a degeneracy of the vacuum of the corresponding classical theory.

[7] Thesis details sphaleron EW. http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/SchaldachJoerg/diss.pdf
Configurations which minimize the potential energy for half-integer NCS are called sphalerons. The word is taken from the Greek language and means "ready to fall", which resembles the fact that it sits on top of the minimal energy curves (slippery). The energy barrier between topologically distinct vacua with adjacent NCS is called "sphaleron barrier". [7]
In this work we considered various aspects of fermion number violation in the electroweak theory. This effect is based on topological properties of the classical SU(2) gauge field. So after formulating the model itself, we discussed the second Chern character q and the Chern–Simons number NCS, which are topological winding numbers of the SU(2) field on spacetime and 3-space, respectively. 260 pages.

Algebraic Topology Spanier: 90 Usually a structure group G is provided for the bundle consisting of homeomorphisms of F.
P 92 A "lifting function" assigns to each point $e \in E$ and path ω in B starting at π(e) a path λ(e,ω) in E starting at e that is a lift of ω. Then the map π is a fibration.

Web Reference:
https://ncatlab.org/nlab/show/electromagnetic+field
(printed out_) Includes Dirac Monopole:
Connections, gauge theory and characteristic classes, thesis:
https://esc.fnwi.uva.nl/thesis/centraal/files/f883485001.pdf
Chern-Simons theory is a gauge theory that does not need a metric; it is therefore called a topological gauge theory. (captured)

We can think of the global wavefunctions as not really functions on M × R, but sections of a possibly non-trivial U(1)-bundle P which we might call the phase bundle, and imagine the fibre as keeping track of the "phase" of the quantum particle. This condition is equivalent to the condition that the curvature F_A has integral periods, $\int F \in 2\pi \mathbb{Z}$.
$F = (1/2)F_{ab}\, dx^a \wedge dx^b$, $F = \mathbf{d}A$; $\mathbf{dd}A = dF = 0$ (*for, MTW Ch 4 forms*), zero exterior derivative.

The Yang-Mills YM curvature form is $F = dA + A\wedge A$. Also, Faraday $F = E\wedge dt + B$.
Monopoles don't really exist, but their math is very rich and of interest by itself .
The simplest generalization of the Dirac monopole is the 't Hooft-Polyakov monopole
The first Chern class of any U(1) bundle over the 2-sphere is an integer,
$(i/2\pi) \int F = -m \in Z$

Zee says. P230 YM $L = (-1/2g^2)\ \mathrm{tr}\ F_{ab} F^{ab}$. Has a chapter on condensed matter and QFT.
p 295 components: $L = Lo + a\,\partial\, a + aJ$ and in forms ada, The CS term has the effect of endowing the charged particles in the theory with flux
Frankel: Chap 22, Chern Forms and Homotopy Groups

For SU(N) instead of U(N) tr $\theta = 0$ so $c_1(E) = 0$ and $C_2(E) = (1/8\pi^2)\mathrm{tr}(\theta\wedge\theta)$ is the 4-form appearing in the winding number of an SU(2) instanton

MTW p198 Cartan invented the exterior derivative in 1901 at the age of32.
Understanding differential geometry should be done at all three levels: pictorial geometry, by components, and by abstract differential geometry, e.g., a tangent vector **A** in its own right or by components $A = A^o e_o + A^1 e_1 + A^2 e_2 + A^3 e_3$.
P 223 connection $\Gamma\ ^a{}_{bg} = -\langle\nabla_g \omega^a, e_b\rangle$, Curvature tensor Riemann $R(A,B) \equiv [\nabla_a, \nabla_b] - \nabla_{[A,B]}$
Curvature 2-forms $R^\mu{}_v \equiv d\omega\ {}^\wedge\mu_v + \omega^\mu{}_\alpha \wedge \omega^\alpha{}_v$ [MTW p 351]

Ellis, ArXiv: Remarkably, the prospective IceCube constraints on **sphaleron**-induced transitions are comparable to those from the LHC, as seen in Fig. 5, with IceCube having an advantage for large sphaleron energies ESph and the LHC at small ESph. The crossover is currently close to the nominal value ESph = 9 TeV.

Notes from March Physics Notes 2016:
Benn/Tucker p 177 Galilean bundle 4d fiber bundle projected to time, each fiber is E^3.
2-form $F = B + dt \wedge E$ (p 180)
In nontrivial topology, one can have a U(1) bundle on S2 topologically classified by an integer (the first Chern class). In physics the resulting field strength is called a Dirac Monopole, and the first Chern class is called the charge of the monopole.
[Gauge Theory ref lost]

other NOTES:

In EWBG, the Universe undergoes a first order phase transition during which electroweak symmetry is broken. The electroweak phase transition (EWPT) proceeds via nucleation of bubbles of broken electroweak symmetry as the Universe cools through a nucleation temperature TN that lies below the phase transition critical temperature, TC. This transition, which satisfies the Sakharov out-of-equilibrium condition, is analogous to the condensation of water droplets from vapor with decreasing temperature. Sakharov's second ingredient is provided by C- and CP-violating interactions of new particles at the bubble walls. These interactions ultimately induce the sphalerons to create baryons that diffuse inside the expanding bubbles where they are captured and protected from being washed out by inverse sphaleron processes.

The LHC and prospective future colliders are well-suited to looking for the particle physics ingredients needed for the first order EWPT. http://arxiv.org/pdf/1604.05324.pdf "The Higgs Portal and Cosmology"

A true Klein bottle cannot exist in 3-D so our usual glass bottle picture is a Kludge.
A Principal fiber bundle is a gauge type. A connection is a gauge potential, EM has connection on a trivial U1 bundle and monopole means connection on a nontrivial U1 bundle. (Topology and Gauge Theory in Physics) "*In 1944 fiber bundle theory became important in topology with Chern's generalization of the Gauss-Bonnet theorem to four dimensions.*
"Gauge Theory" EM and U1 "the vertical automorphisms of the bundle will reproduce the gauge transformation of EM."
Hurewicz in 1935 developed the concept and theory of higher dimensional homotopy groups." (more?)
[12] Def: A local section of E is a smooth map s from a neighborhood U in M to E such that $\pi \circ s(x) = id(x)$ over all $x \in U$ (ie the image of x lies in the fibre $\pi^{-1}(x)$.
Def: A **lift** of a smooth path $\gamma_M:[0,T] \rightarrow M$ in M is a smooth path $\gamma_E:[0,T] \rightarrow E$ in E such that $\pi \circ \gamma_E = \gamma_M$. (this actually makes clear sense)

Locally trivial means diffeomorphic to R^n. *The cylinder C ~ S1×[a,b] global product is a trivial bundle with fiber F = [a,b]⊂ R.* But, the Möbius strip, $Mö^2$, is not a trivial bundle since it looks only locally like $S^1 \times [a,b]$ for open subsets U⊂ M.
A bundle or fiber bundle is trivial if it is isomorphic to the cross product of the base space and a fiber.
We say that a manifold is parallelizable if its tangent bundle is trivial.
Chern: Ask how many different bundles are there over M and how many are non-trivial.

Characteristic classes are the basic cohomological invariants of bundles and have a wide variety of applications throughout topology and geometry. Characteristic classes were introduced originally by E. Stiefel in Switzerland and H. Whitney in the United States in the mid 1930's. [13, Cohen]…. In the early 1940's, L. Pontrjagin, in Moscow, introduced new characteristic classes by studying the Grassmannian manifolds, using work of C. Ehresmann from Switzerland. In the mid 1940's, after just arriving in Princeton from China, S.S Chern defined characteristic classes for complex vector bundles using differential forms and his calculations led a great clarification of the theory.

U(1): The Hopf fibration is an example of a non-trivial circle bundle.
Examples of non-trivial fiber bundles include the Möbius strip and Klein bottle, as well as **nontrivial covering spaces.** WIK
The unitary group U(n) has universal cover SU(n) × ℝ.
The n-sphere Sn is a double cover of real projective space RPn and is a universal cover for n > 1.
covering maps of topological spaces, using the classic example of the real line winding onto to the circle.
the unit quaternions double cover SO3.
So it's a Moebius band, as you say. The Moebius band is a quotient of a cylinder, which is a quotient of the real plane.
<u>Perhaps the simplest example of a nontrivial bundle E is the Möbius strip.</u>

http://arxiv.org/pdf/hep-th/0611201.pdf lecture notes, Chern.

Topology of Fibre bundles and Global Aspects of Gauge Theories, Andr´es Collinucci
"the non-triviality of the Mobius strip had to do with the fact that one could not find a global trivialization. We now understand that this is the case because one cannot define a linearly independent (which in one dimension means everywhere nonzero) section on Mo … A principal bundle is trivial if and only if it allows a global section. (i.e., a Mobius rectangle with opposite corners identified has a section curve that has to cross zero).
Figure 3: The principal bundle P (S1, Z2) associated to the Mo˙bius strip is a double cover of the circle. (two circular curves going around without intersecting)

NICE EXAMPLES:
"Instantons are traditionally defined as smooth finite action solutions of Yang-Mills theory on 4-dimensional Euclidian space R4. We will only consider the case of SU(2). There exist no non-trivial bundles over R4, but the finiteness of the action imposes boundary conditions at infinity, which allow for the existence of topologically non-trivial solutions of the field equations. PAGE 33 is interesting: "To gain more control over the situation and allow for a bundle description of instantons, we consider a one-point compactification of R4 to S4, by adding to it the point at infinity, $R^4 \cup \{\infty\} = S^4$. This means that we want to look at principal SU(2)-bundles over S^4, P(SU(2),S^4).
And it talks about "sufficiently large" 3-spheres: S_∞^3.
the sphere S^2 is its own universal cover.
the real line covers the circle: (i.e., an infinite helix above a circle covers the circle).

Lecture XI- **Homotopies** of Maps. Deformation retracts:
http://nptel.ac.in/courses/111101002/downloads/lecture11.pdf
Georgi, Glashow SU(5) was one guide to baryon number violation (baryogenesis). And SO(10) is a guide model to lepton number violation (lepto-genesis). However, experimental proof of lepto-genesis requires establishing the Majorana nature of the ordinary neutrino.
Dn is $|x| \leq 1$ (closed). Homotopy is an equivalence relation. Homotopy of paths is generalized to homotopy of a pair of continuous maps between topological spaces. Homotopy has proved to be the most important notion in topology…" Think of spheres as being UNIT spheres, then R^n-{0} $\rightarrow S^{n-1}$ is r(x) = x/||x||.

It is often difficult in topology to prove things up to homeomorphism. Often, we only prove stuff up to homotopy. In fact, much of algebraic topology classifies topological spaces up to homotopy.
If a space has the homotopy type of a point, it is called contractible.

For a Yang–Mills theory these inequivalent sectors can be (in an appropriate gauge) classified by the third homotopy group of SU(2) (whose group manifold is the 3-sphere S^3). A certain topological vacuum (a "sector" of the true vacuum) is labelled by an unaltered transform, the Pontryagin index. As the third homotopy group of S^3 has been found to be the set of integers, Z. WIK Instanton.

Wik: Trace of a product[edit]
The trace of a product can be rewritten as the sum of entry-wise products of elements:
\operatorname{tr}(X^{\mathrm T}Y) = \operatorname{tr}(XY^{\mathrm T}) = \sum_{i,j}X_{ij}Y_{ij}.
This means that the trace of a product of matrices functions similarly to a dot product of vectors. For this reason, generalizations of vector operations to matrices (e.g. in matrix calculus and statistics) often involve a trace of matrix products.

Trace of matrices: Tr(AB) = Tr(BA).

In YM "we may consider the vacuum state in which field strength F or θ vanishes…. In the AB effect we have seen that a parallel translation about S1 does not return a vector to itself, in spite of the fact that the connection is flat. (??) A has more information than vanishing field strength, and flat connection has more information than 0 curvature alone.
Frankel p 558: "In the 4-dimensional Yang-Mills case (with G = SU(2)) there will be an infinity of inequivalent vacua, each one characterized by the degree or "Winding number" of the map g: S3→ SU(2)…"
585 Chern suggested the possibility of expressing winding number in terms of an integral of a 4-form involving curvature.—the differential of a 3-form, the Chern-Simons 3-form. The winding number of the instanton is - ∫ c2 over R4 for SU(2) bundles. P 609.

http://www.scholarpedia.org/article/Axial_anomaly
The axial anomaly is a quantum term that violates the classical conservation of the axial current…. The physical interpretation of instantons is that they provide a semi-classical signal for the occurrence of quantum tunneling; here it is the tunneling between homotopy classes of gauge fields. CS also mentioned.

Instantons in QCD 9610451, 140 pages Schafer
QCD Instantons play a role in chiral symmetry breaking. They are localized (e.g., 1/3 fm) regions of space-time with very strong gluonic fields. The instanton solution of the YM equations was discovered by Polyakov, 1975 and later attributed to a tunneling event between degenerate classical vacua. The term instanton was coined by 't Hooft.
Instantons are topologically nontrivial solutions of Yang–Mills equations that absolutely minimize the energy functional within their topological type.
Instantons play a central role in the nonperturbative dynamics of gauge theories.

All global structure in field theory is controlled by fiber bundles. Soliton solutions such as instantons and monopoles are classified according to characteristic classes of fiber bundles."
Washington: 1.1 The product formula in action and Chern classes as ob- structions to "global generation"
(differential geometry) Given a smooth closed curve C on a surface M, and picking any point P on that curve, the **holonomy** of C in M is the angle by which some vector turns as it is parallel transported along the curve C from point P all the way around and back to point P.
YM instanton and QCD instanton are nontrivial class of the principal bundle underlying the YM gauge field.
Manifolds can describe translational degrees of freedom and Fiber bundles can describe internal degrees of freedom (such as spin and isospin)
U(1) gauge potential $A = A_i dx^i$ is a connection on a complex line bundle $R^3 \times C$. A section of the line bundle gives a complex value function ψ, and the covariant derivative on the line bundle is $D\psi = d\psi - iA\psi$
Charles Ehresmann (1950) was a student of Cartan and thought of a connection in a principal bundle as a specification of horizontal and vertical vector fields. A parallel translation is a lifting of a curve from B to a curve in P which is horizontal.
A connection says how to transport data along a curve

The main problem in topology is classifying topological spaces up to homeomorphisms (homotopy type).

In the mathematical field of topology, the **Hopf fibration** (also known as the Hopf bundle or Hopf map) describes a 3-sphere (a hypersphere in four-dimensional space) in terms of circles and an ordinary sphere. Discovered by Heinz Hopf in 1931, it is an influential early example of a fiber bundle. Technically, Hopf found a many-to-one continuous function (or "map") from the 3-sphere onto the 2-sphere such that each distinct point of the 2-sphere comes from a distinct circle of the 3-sphere (Hopf 1931). Thus the 3-sphere is composed of fibers, where each fiber is a circle — one for each point of the 2-sphere.
S3 is not globally a product of S2 and S1 although locally it is indistinguishable from it.
It applies to S1, S3, S7, and S15, but usually S2→ S2 with fiber S1.

In geometry, Villarceau circles /viːlɑːrˈsoʊ/ are a pair of circles produced by cutting a torus obliquely through the center at a special angle. Given an arbitrary point on a torus, four circles can be drawn through it. One is in the plane (containing the point) parallel to the equatorial plane of the torus. Another is perpendicular to it. The other two are Villarceau circles. They are named after the French astronomer and mathematician Yvon Villarceau (1813–1883).

Burke 159. Exterior calculus can be extended to n-d metric spaces using the Hodge star $*:\Lambda^{p}\rightarrow \Lambda^{n-p}$. There is also a sharp operator mapping 1-form α → tangent vector #α .
$\#dx = \partial_x$. $*1 = dxdy$. In Minkowski 2-space, $\#dt = -\partial_t$.
Laplaces eqn in E3 has α = pdydz + qdzdx + rdxdy, so *α = pdx + qdy + rdz

P 86 π: E→ B, a particular field is a section given by a reverse function Γ:B→ E with π∘ Γ(b) = b.
P 88 A simple example of a cotangent bundle is taking gradient phi for every point of M.

A section of a fiber bundle gives an element of the fiber over every point in B. Usually it is described as a map s:B->E such that pi degrees is the identity on B. A real-valued function on a manifold M is a section of the trivial line bundle M×R. Another common example is a vector field, which is a section of the tangent bundle. Wolfram
This sheaf is called the sheaf of sections of f, and it is especially important when f is the projection of a fiber bundle onto its base space. Notice that if the image of f does not contain U, then Γ(Y/X)(U) is empty. For a concrete example, take X = C \ {0}, Y = C, and f(z) = exp(z). Γ(Y/X)(U) is the set of branches of the logarithm on U.

An index measures difference in the number of “zero modes” or differences in dimensionality of two spaces. Fei Han said the Chern character is given by the map that “crosses with the circle.” Using K-theory and de Rahm cohomology. (K for Klasse, class).

http://arxiv.org/pdf/1605.08081.pdf
The Abelian anomaly is responsible for the decay $\pi_0 \rightarrow \gamma\ \gamma$. It is represented by the triangle diagram with two vector current vertices that couple to the two photons and one axial vertex linking to the π_0 . The anomaly is related to the Atiyah-Singer index theorem in topology…. In 1949, Steinberger [53] had already calculated in his PhD a Feynman diagram, a triangle diagram with two vector current vertices and one axial vertex.

<u>Tony Zee QFT:</u> p219 Components of dF = ddA = 0 imply the Bianchi identity for EM. Monopoles by two caps –N and –S, there is no dϕ there, and separate A's give difference at equator p 221 imply g = 2π n/e
p223 string theory has numerous p-forms.
295 Chern-Simons term is "ada" = $\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a$. The CS action is gauge invariant. The CS term has the effect of endowing the charged particles in the theory with flux. $S = \gamma \int_M d^3x\, ada$ is topological. The effective theory of the Hall fluid turns out to be a Chern-Simons theory."
One form $A = A_\mu dx^\mu$ (subscript for form).

<u>MORE MATH:</u>

Burke 159. Exterior calculus can be extended to n-d metric spaces using the Hodge star $*:\Lambda^p \rightarrow \Lambda^{n-p}$. There is also a **<u>sharp</u>** operator mapping 1-form α → tangent vector #α . $\#dx = \partial_x$. *1 = dxdy. In Minkowski 2-space, $\#dt = -\partial_t$.
Laplaces eqn in E3 has α = pdydz + qdzdx + rdxdy, so *α = pdx + qdy + rdz

P 86 π: E→ B, a particular field is a **section** given by a reverse function Γ:B→ E with π∘ Γ(b) = b.
P 88 A simple example of a cotangent bundle is taking gradient phi for every point of M.

A section of a fiber bundle gives an element of the fiber over every point in B. Usually it is described as a map s:B→ E such that pi degrees is the identity on B. A real-valued function on a manifold M is a section of the trivial line bundle M×R. Another common example is a vector field, which is a section of the tangent bundle. Wolfram

The Atiyah–Singer theorem was announced by Atiyah & Singer (1963).
In differential geometry, the jet bundle is a certain construction that makes a new smooth fiber bundle out of a given smooth fiber bundle. It makes it possible to write differential equations on sections of a fiber bundle in an invariant form. Jets may also be seen as the coordinate free versions of Taylor expansions.

Historically, jet bundles are attributed to Ehresmann, and were an advance on the method (prolongation) of Élie Cartan, of dealing geometrically with higher derivatives, by imposing differential form conditions on newly introduced formal variables.
For $p \in M$, let $\Gamma(\pi)$ denote the set of all local sections whose domain contains p.
the Chern character maps K-theory (vector bundles) to cohomology (differential forms).
The basic strategy of the local argument, as simplified by Getzler, is to invent a symbolic calculus for the Dirac operator which reduces the theorem to a computation with a specific example. This example is a version of the quantum-mechanical harmonic oscillator
http://mathoverflow.net/questions/23409/intuitive-explanation-for-the-atiyah-singer-index-theorem
the main idea is the Bott periodicity theorem.

http://isites.harvard.edu/fs/docs/icb.topic1146666.files/IV-6-Anomalies.pdf
Anomalies, Swartz, 2013. Looks interesting.

Kaku QFT p 414 “An anomaly is the failure of a classical symmetry to survive the process of quantization and regularization.” (such as expecting axial currents to be conserved in a chiral gauge theory)
The decay of the pi zero was not occurring at the expected rate and needed to include an anomaly. An internal triangular fermion loop of u quarks or d quarks (or antiquarks), the triangle anomaly spoils the renormalisability of the SM.
Weinberg II p 361 Standard theory would predict a pi zero decay rate near 1.9E+13 /sec but actual rate 4.4E+16/sec.
the currently accepted value of $\tau(\pi^o)$ is 0.8 10^-16 s. (2013)
the charged pions π+ and π− decaying with a mean lifetime of 26 nanoseconds (2.6×10−8 seconds), and the neutral pion π0 decaying with a much shorter lifetime of 8.4×10−17 seconds. Charged pions most often decay into muons and muon neutrinos, while neutral pions generally decay into gamma rays.
The π0 meson has a mass of 135.0 MeV/c2 and a mean lifetime of 8.4×10−17 s. It decays via the electromagnetic force, which explains why its mean lifetime is much smaller than that of the charged pion (which can only decay via the weak force). WIK

A standard example of an anomaly is the (anomalous) decay time of the neutral pion. Charged pi-mesons have a lifetime of 26 nanoseconds and decay via the weak force. The pi-zero with mass 135 MeV decays electromagnetically and only lives for 8.4x10-17 seconds (about a thousand times faster than older theory predicted). The cause of this is a Feynman diagram with a closed triangular quark flow between an input π o vertex and 2 photon output vertices. Roman Jackiw used an older Steinburger idea of a proton triangular current but now using QCD. He found that his anomaly used the Atiyah-Singer index theorem which in turn used a Chern character (proportional to trace(FF)).

1. Chiral Anomaly
In classical physics there is said to be a symmetry when the action S(ϕ) is invariant under the transformation $\phi \rightarrow \phi + \delta \phi$, while in quantum mechanics the path integral $\int$ Dψ eiS(ψ) must be invariant for a symmetry to be present. The transformation from classical to quantum mechanics does not always retain a given symmetry. Otherwise said: Symmetries in terms of classical, commuting variables may not be retained when expressed in terms on non-commuting quantum variables. Such a symmetry is said to have a “quantum symmetry anomaly”. http://arxiv.org/pdf/1605.09214.pdf

WIK Homology was originally a rigorous mathematical method for defining and categorizing holes in a manifold. Loosely speaking, a cycle is a closed submanifold, a boundary is the boundary of a submanifold with boundary, and a homology class (which represents a hole) is an equivalence class of cycles modulo boundaries….. A 0-dimensional hole is simply a gap between two components, consequently
HoX describes the path-connected components of X.
Notice that, algebraically, we define a hole to be a cycle that does not bound, i.e., we say that the homology is non-trivial , or that there is an n-hole if the quotient Zn/Bn$\neq$id. If you look, e.g., at the case of a 2-torus $T^2=S^1\times S^1$,
you will see that, e.g., a meridian is a cycle that does not bound, because its removal will not disconnect the space. Similarly for any strictly latitudinal curve. These two cycles (simple-closed curves in the space) generate the homology of the torus.

A torus has one connected component b_o, two circular holes b1 (the one in the center and the one in the middle of the donut), and one two dimensional void (b2 , the inside of the donut—the number of voids or cavities) yielding Betti numbers of 1,2,1 or Poincare' polynomial 1 + 2x + x2.

1605.09433 Hopf fibration. In mathematics, the Hopf fibration describes S3 in terms of a disjoint union of circles S1 and an ordinary S2 with fiber structure S1→ S3 → pi S2. With pi for projection . S3 can be C2 with z0^2 + z1^2=1

http://arxiv.org/pdf/1605.08081.pdf
The Abelian anomaly is responsible for the decay $\pi 0 \rightarrow \gamma\ \gamma$. It is represented by the triangle diagram with two vector current vertices that couple to the two photons and one axial vertex linking to the $\pi 0$. The anomaly is related to the Atiyah-Singer index theorem in topology…. In 1949, Steinberger [53] had already calculated in his PhD a Feynman diagram, a triangle diagram with two vector current vertices and one axial vertex.

WIK In mathematics, the Bott periodicity theorem describes a periodicity in the homotopy groups of classical groups, discovered by Raoul Bott (1957, 1959), which proved to be of foundational significance for much further research, in particular in K-theory of stable complex vector bundles, as well as the stable homotopy groups of spheres. K Theory: The subject can be said to begin with Alexander **Grothendieck** (1957), who used it to formulate his Grothendieck–Riemann–Roch theorem. It takes its name from the German **Klasse**, meaning "class". Grothendieck needed to work with coherent sheaves on an algebraic variety X.
Kernel: In mathematics, and more specifically in linear algebra and functional analysis, the kernel (also known as null space or nullspace) of a linear map L : V → W between two vector spaces V and W, is the set of all elements v of V for which L(v) = 0, where 0 denotes the zero vector in W. That is, in set-builder notation, ker(L)={v∈ V|L(v)=0}.

There are large-Chern-number topological phases, Floquet topological phases, Dirac cones, Band topology can be characterized by Chern numbers (e.g., ± 7?? Minus?) 1601.04437, There are also Chern Insulators??
"A Chern insulator is a zero magnetic field version of the quantum Hall effect (QHE)." Topological properties emerge from the band structure, and at least one band is a non-zero Chern number, c." And c can be larger than one.
Rutgers "In order to have a robust non-zero Chern number, a system must have broken time reversal symmetry and strong spin-orbit coupling.
Topological insulators have insulating bulk and conducting edge or surface states immune to small perturbations.

Connectedness and dimension are invariants.
Planet math: A topological invariant of a space
If X is a property that depends only on the topology of the space, i.e. it is shared by any topological space homeomorphic to X.

Properties of a space depending on an extra structure such as a metric (i.e. volume, curvature, symplectic invariants) typically are not topological invariants, though sometimes there are useful interpretations of topological invariants which seem to depend on extra information like a metric (for example, the Gauss-Bonnet theorem).

For a field theory with Chern-Simons action, expectation values of Wilson line operators are topological invariants.

J. W. Alexander, Topological Invariants of Knots and Links, 1927 They are related to the knot group of Dehn.
A topological invariant is any property of a topological space that is invariant under homeomorphisms. E. g. Connectedness,
If two spaces are not homeomorphic, it is sufficilent to find a topological property which is not shared by them
The existence of anyons comes from considering topological differences between paths in 2 and 3 d space. "In 2-d space, we cannot deform any arbitrary path into our initial exchange path." Anyons have been detected.
Geometric topology revolves around manifolds and embeddings of them (and foliations which are nice ways for slicing up manifolds, surgery theory..).
Topological space has a notion of nearness and the preservation of points being near each other.
Topology is (very roughly) the study of shapes that can be stretched, squished and otherwise tortured while keeping near points together.
One of the ways people found to deal with those difficulties is to create gadgets (officially called functors) that map topological spaces into objects that are easier to handle - algebraic objects like vector spaces and groups.

Today, the applications of topology in physics are numerous [27]. A large aspect of this is the recent revolution called "topological matter" such as topological insulators, topological phases, topological superconductors and topological semimetals. The importance of topology emerged with the recognized importance of gauge theories. These topological phases are characterized by "topological invariants that have a global dependence on characteristic parameters of the system." Continuous deformation of one will not extend to those of another. Quantum effects are usually low energy modes and the topology of the bands of energy spectrum.
[Italics means that the concept has already been used in the Paper].
There is a Bloch bundle E(T) over the Brillouin zone T where the fibres are the spaces of states with the same Bloch momentum k [27]. For the integer quantum Hall effect, the Bloch bundle $E(T^2)$ is over the 2D Brillouin torus.
Perelman proved the Poincare conjecture that any 3D topological space X with $\pi_1(X)=0$ is topologically equivalent to the S^3 sphere [27].

Covariant Derivative Issues

Dave Peterson, 5/20/17 – 9/1/17

… [Einstein 1912] "I suddenly realized that Gauss's theory of surfaces holds the key for unlocking this mystery." … Then, I told Grossmann that I "needed a geometry which allowed for the most general transformations that leave the metric invariant [$ds^2=g_{\mu\nu}\, dx^\mu\, dx^\nu$]. Grossmann replied that Einstein was looking for Riemannian geometry" [16, p213, Pais].

In curved spaces, the idea of covariant differentiation includes "connections" expressing "parallel displacements" of vectors and vector bases. For tangibility and clarity, these are calculated here in a variety of ways for the simple special case of translations along the latitude of a 2-sphere. Connections expressed as Christoffel symbols versus 1-forms are related by "scale factors" h_i. The language of differential forms provides an economy of expression facilitating understanding for 2d surfaces and higher dimensions. An ultimate goal is understanding "Curvature." But differential geometry uses a large variety of notations that make understanding difficult.

"We shall not cease from exploration. And the end of all our exploring
Will be to arrive where we started And know the place for the first time." T.S.Elliot.

After many years of studying general relativity [GR], I am finally taking a fresh look at its basic differential geometry foundations. As "old students" from the '60's and '70's, we had gotten used to performing "index manipulation" of tensors; but more recently general relativity has evolved towards greater use of abstract differential geometry stressing the language of differential forms. Both the old and new ways are shown in the big standard text by "Misner-Thorne-Wheeler" [6, or "MTW"] . The next challenge for us now is to try to make some sense of the "totally modern" GR text by Wald [15] [and see Figure 1 below].

*{Example: {$de_\sigma = e_\mu \wedge \omega_\sigma{}^\mu$, $R_{abc}{}^d\, w_d = (\nabla_a \nabla_b - \nabla_b \nabla_a) w_c$, $\boldsymbol{R}_\mu{}^\nu = d\omega_\mu{}^\nu + \omega_\mu{}^a \wedge \omega_a{}^\nu$ }. **Eqn. 1**.*

The math itself is really not new, the explicit and abstract approaches both go back a century. It is a hope that understanding the differential geometry currently being used for general relativity space-time may be eased first by an adequate study of "elementary" differential geometry for simpler Euclidean spaces like E^3 [1].

A key concept in both flat and curved spaces is the use of the **"covariant derivative"--** a way of specifying a derivative along tangent vectors" of a surface. This concept is a generalization of the "directional derivative" of a scalar function f(x,y,z) from vector calculus, $\nabla_v f = (\mathbf{v}\cdot \nabla f)$ -- the projection of a gradient onto a tangent vector v to a surface or curve. A primary virtue of covariant differentiation is that it converts tensors into other tensors; it preserves their invariance under coordinate transformations. An ordinary derivative usually lacks this property. For flat space-time and Cartesian coordinates, the covariant derivative then reverts back to the ordinary derivative. The concept of covariant derivative goes back to Ricci and Levi-Civita (~ 1901 and earlier) – in time for application by Einstein and Grossman.

The concept of directional derivatives also applies to vector fields $\nabla_v \mathbf{W}$ where $\mathbf{W} = w_1e_1+w_2e_2+w_3e_3 = w_i(x,y,z)e_i$ where the gradient of each function coefficient, w_i, is separately projected onto the same tangent vector v (and the "elementary" e_i's may be unit vector bases). $\nabla_v W$ is the rate of change of W in the v(p) direction ("tangent" vectors have individual points of origin, p). The covariant derivative of vectors must allow for correction terms due to the possible "rotation" of basis vectors on curved surfaces, and the added terms are called "connections."

Several different notations for expressing directional derivatives of a scalar function, f, include:

$\nabla_v f = \partial_v f = \langle df,v\rangle = v[f] = \mathbf{v}\cdot(\nabla f) = (v\cdot\nabla)f = (v^i \partial / \partial_{xi}) f = v^i \partial_i f$. **Eqn. 2.**

Or, $v[f] = (d/dt)(f(p+tv))\,|_o$ as a point, p, advances with time in the v direction. And for vector fields, $\nabla_v W = W(p+tv)\,'(o)$ measures the rate of change of W(p) as p moves in the v direction evaluated at t = 0 [1].

Stumbling blocks or early learning problems arise in using covariant derivatives:

One is due to the variety and inconsistency [e.g., Eqn. 1 above] in labeling notation and conventions from text to text. I think that a huge problem is the use of the symbol "**e**" for both unit vectors (orthonormal bases) and non-unit coordinate vectors. Then the coordinate values have to be magnified or reduced in magnitude to compensate as the base vector is lowered or raised from unit length {For metric spaces, see the use of h_μ scale factors below}. The popular text by Misner, Thorne and Wheeler (MTW [6]) uses both but is careful to place little hats (^) on unit vectors, but generally one has to know context to see if e is unit or not. The same applies to dual basis vectors—they may be labeled as θ^i or ϕ^i or ω^i or σ^i – but you may not immediately know if they derive from unit vectors bases or not.

Introductory geometry in the usual Euclidean space E^3 might only use orthonormal unit vector bases (e.g., O'Neil [1]) while tensor application would mainly use covariant and contravariant bases which are **not** usually unit vectors. General relativity most commonly uses "coordinate bases." So, we have to be able to transform between "coordinate bases" (like ∂_μ's) and "associated orthonormal non-coordinate bases" (like unit vectors). There is a fr**ee choice** between using coordinate bases and orthonormal bases. Orthonormal bases are best for measurements and physical interpretation. This problem with having orthonormal unit bases or not carries over to Christoffel symbols (Γ 's) and Riemann tensors. There are two different types and they have different values; and the convention being used may not be clear up front.

A later problem is translating index conventions from "only low" (e.g., turning connection ω_{ij} with "orthonormal" bases) to covariant/contravariant indices (e.g., $\omega^i_{\ j}$). For unit frame bases, these are the same, but contravariant bases are not of unit length. There are also ω's where the i is directly above the j rather than staggered sideways. For the important "curvature 2-form" Ω_{ij} also = $\Omega^i_{\ j}$ when orthonormal bases are used. Understanding a field of study requires dovetailing together a variety of approaches over many published papers and multiple texts, and this can be a challenge.

We are all familiar with the idea that "vectors" in Euclidean space only have a direction and magnitude and can be freely transported anywhere we wish to move them (e.g., the "parallelogram rule" for vector addition, A+B = B+A). But that is not true in curved spaces or surfaces (e.g., the sphere or the saddle shaped surfaces). Vectors are first replaced with "tangent vectors" on a surface depending on individual points of application – a vector from a given point p to another point. The important concept of parallel transport of these vectors on a curved surface often results in rotations of the basis vectors. The degree of relative rotation often might initially appear to vary with different methods of calculation. We must verify that all approaches are consistent. Base rotation "connections" may be expressed as Γ symbols ("Christoffel symbols") or alternatively by ω's as "1-forms." Understanding these somewhat difficult terms may require tangible calculations for simple intuitive cases beyond just metric index manipulations.

Initially, we take directional derivatives with respect to given vectors at a point, v_p. But a little later on we also take them with respect to **indices**, like $\nabla_k \mathbf{A} = \mathbf{A}^i{}_{;k}$ using a semicolon for covariant derivative (using just a comma "," means taking just ordinary derivatives). Integer indices refer to the labeling of basis vectors. Note that affine connections are **defined** in terms of the turning of basis vectors towards each other: $\nabla_\alpha X_i = \Gamma^j{}_{i\alpha} X_j$ [5]. With these corrections for covariant derivative, we get an invariant tensor. The other more abstract approach uses differential "1-forms" ω and unit "frame" fields, E_i, $\nabla_v E_i = \Sigma\, \omega_{ij}(v) E_j$ [1] . "Each geometric surface has its own notion of covariant derivative" [1, p 338], but every chosen metric $g_{\mu\nu}$ defines a unique Γ connection. "Covariant differentiation is completely determined by its action on a basis." And then after knowing that, we can apply it to 3-d vector fields, W, on the separate component functions of W.

A vector ***A*** can be written in three basic ways depending on the choice of basis vectors. $$A = a_\mu\, e_\mu = A^\mu \partial_\mu = A_\mu\, e^\mu \qquad \textbf{Eqn. 3.}$$

(the usual "Einstein convention" says that repeating indices means: Σ "sum over" index values μ = 1,2,3…). Then we can dispense with any need to show the Σ_i's. The components in the first case, $\mathbf{A} = a_\mu e_\mu$, are called "physical" with orthonormal bases such as i,j, k (unit vectors). This is typical in standard Euclidean vector analysis, and the component magnitudes are directly useful. **Unit** basis vectors are basis direction vectors that are **normalized** to length one as $e_\mu = (\partial r/\partial x^\mu)/\|\partial r/\partial x^\mu\|$ -- a carrot ^ over the e might be preferred for clarity.

The next term for **A** is called "contra-variant" with upper index A^μ and vector bases that are tangent vectors: $\partial \mathbf{r}/\partial x^\mu = \partial_\mu \boldsymbol{r}$ [conventionally shortened to just $\partial/\partial x^\mu = \partial_\mu$ *and without the location vector* ***r*** *;* and we should perhaps try to avoid using the symbol e_μ for these non-unit bases and restrict that to unit vectors]. The danger of mistaking the contravariant base from the general act of taking partial derivatives could be avoided by using a different symbol like α_μ (alpha) instead of ∂_μ as a basis vector (but that is rarely done). An example of this "coordinate" base is $\partial_\phi = r \sin\theta\, e_\phi$ for spherical coordinates. We hold r and θ constant and ask how much real spatial distance results from a change in angle dϕ.
Once you know the bases, the metric tensor is merely the product of bases, $g_{\mu\nu} = \partial_\mu \cdot \partial_\nu = e_\mu \cdot e_\nu$ (if the e's are coordinate bases).

The last form for $\mathbf{A} = A_\mu e^\mu$ is "co-variant" using a lower index and bases that are gradients of the coordinate curves, $e^\mu = \nabla u_\mu = \beta^\mu = \mathbf{d}x^\mu$ -- and again not unit vectors (in the "pretty" language of "forms," the **d** here is called "exterior derivative" and is no longer thought of as a "tiny" Δ – more discussion later on). The non-unit (non-physical) contravariant and covariant bases are the ones used for "conventional index gymnastics computations." Using e's here for co-vector bases is ok because they have superscripts and avoid confusion.

The relationship between co- and contra- bases is $\mathbf{d}x^\mu(\partial/\partial x^\nu) = \delta_{\mu\nu}$ (i.e., 1 or 0).

Or,

$e^i \cdot e_j = \nabla u^i \cdot (\partial \mathbf{r}/\partial u^j) = (\partial u^i/\partial x)(\partial x/\partial uj) + \ldots = \partial u^i/\partial u^j = \delta^i_j$. As a particular example for 2d polar coordinates, (r,θ), position $\mathbf{r} = r\cos\theta\, i + r\sin\theta\, j$ where $\theta = \mathrm{Tan}^{-1}(y/x)$.

$d\mathrm{Tan}^{-1}x = 1dx/(1+x^2)$, so $\nabla\theta = -iy/r + jx/r$, and $\partial r/\partial\theta = -r\sin\theta i + r\cos\theta j = -yi + xj$.

$$\text{So, } \nabla\theta \cdot (\partial r/\partial\theta) = (y^2+x^2)/(x^2+y^2) = \delta^\theta_\theta = 1.$$

For a Euclidean point, p, a position vector may be given as $\mathbf{r} = xi+yj+zk$. For curvilinear coordinates, we may rename $r = r(u_1,u_2,u_3)$ {for example: "Geographic" coordinates, $\varrho,\vartheta,\varphi$ with angles for longitude and latitude from the equator}. So

$$\mathbf{dr} = (\partial \mathbf{r}/\partial u_1)du_1 + (\partial \mathbf{r}/\partial u_2)du_2 + (\partial \mathbf{r}/\partial u_3)du_3 = (\partial \mathbf{r}/\partial u_\mu)du_\mu \qquad \underline{\textbf{Eqn. 4.}}$$

The coefficients of du_i's in parentheses are called **"scale factors,"** h_1, h_2, h_3 (times corresponding unit vector bases, e_μ).

Then a diagonal metric line element is $ds^2 = \mathbf{dr} \cdot \mathbf{dr} = g_{\mu\mu} du_\mu{}^2 = h_\mu{}^2 du_\mu{}^2$. The tangent bases are $\partial_\mu = h_\mu e_\mu$ while covariant bases are $e^\mu = \nabla u_\mu = e_\mu/h_\mu$.

The covariant base operating on the contravariant tangent base gives:

$dx^\mu(\partial/\partial x^\mu) = 1 = (\nabla u_\mu = e_\mu/h_\mu)(\partial_\mu = h_\mu e_\mu)$ – the h's cancel out.

The conversion from a (contravariant) coordinate vector field $A = A^\mu \partial_\mu$ to a covariant 1-form field (co-vector) $A_\mu dx^\mu$ is via the metric: $g_{\mu\mu}A^\mu = A_\mu$. So given any vector field, we <u>can find an equivalent 1-form</u> using this equation. For 4-vectors and a Minkowski metric (-+++) with index labels $\mu = 0,1,2,3$, the prefix signs of all the scalar functions A^μ will be positive. But the co-vector will change its scalar function for time from + to - .

For "spherical <u>polar</u> coordinates" (r,θ,ϕ), we have the polar angle theta opening down from the north pole of a sphere. A scale factor example here could be $h_\phi = r\sin\theta = \sqrt{g_{\phi\phi}}$, and contra-variant tangent basis is then $\partial_\phi = h_\phi e_\phi$. To compensate for the bases scaling up or down with h, we also need to have $A_\mu = h_\mu h_\mu A^\mu = g_{\mu\mu} A^\mu$ (--usually shown up front in GR texts as a definition for the lowering of an upper index to a lower index using the metric tensor $g_{\mu\nu}$). The change from non-unit base vectors to unit vectors using scale factors is not often stressed or even given in texts.

<u>Special Test Case: A Latitude on a Sphere, S^2:</u>

Some students (like me) desire concrete examples for clarity. An elementary consideration here is the 2-sphere. One part of covariant derivatives is the gradient on S^2 (with constant radius, r, i.e., a 2d-surface case for θ and ϕ). So, we should first know that the gradient of a scalar function in elementary vector calculus, f, is

$$\nabla f = e_\theta(\partial f/\partial\theta)/r + e_\phi(\partial f/\partial\phi)/r\sin\theta \;=\; e_\theta(\partial f/\partial\theta)/h_\theta + e_\phi(\partial f/\partial\phi)/h_\phi . \quad \textbf{Eqn. 5}$$

*[but technically, the e's here should be unit **dual** frame bases θ's or σ's, and the ∇f should be the exterior derivative **df**].*

And for a vector field W on S^2, $\nabla_v W$ uses gradients multiple times— one for each of the function coefficients of the three bases directions.

For Understanding of "parallel transport" of vectors and bases on a sphere: consider movement along a latitude of a sphere or globe, S^2 (at polar θ = some fixed θ_o < π/2). We want to see the resulting rotation of basis vectors from transport in the increasing ϕ direction. Avoid the case θ = π/2 here because it corresponds to the equator and is a "geodesic" or "great circle" without any needed connection term for basis rotations. The goal of the following exercises is to show consistency from a variety of approaches:

Case 1, Visually : To begin, the easiest picture approach to vector rotation from transport is to **draw a circle (radius a) on a flat piece of paper** and draw a set of identical parallel arrow-vectors all pointing in the same direction around that circumference [4]. It is apparent that the angle of these parallel arrows will rotate 2π =360˚ with respect to the circle circumference curve after full motion about the circle. Using scissors, cut to the center in two locations so that a **wedge** of the circle is now missing (wedge angle α). And then fold the paper into a **cone** positioned on the sphere like a dunces cap on a ball. That cone is now tangent to the sphere at the relevant latitude.

If you look from the side (the normals to the cone at the latitude circumference) and rotate the cone-circle of vectors, you will see two things: With increasing angle phi, the vectors seem to rotate clockwise about the circumference while their theta and phi pointing base vectors seem to rotate counter-clockwise to the same compensating degree. When you get to the place where the wedge was missing, the vectors take a jump (a deficit angle) of α =2π (1-cosθ) where theta is the polar angle of a latitude. Let's show that:

Let the slant height of the cone be "a", circumference of the original flat circle is C_a = 2πa, sphere radius is R, radius of latitude circle is r, and the circumference of the cone is C_r = 2πr. The wedge angle is then C_a-C_r = aα = 2π (a-r). But r=Rsinθ , tanθ = sinθ/cosθ = a/R, so cone circumference is 2πr = 2π Rsinθ =2π acosθ .
Then **α =** 2πa(1-cosθ)/a = **2π(1-cosθ**).

A real example of this might be a pendulum bob initially swinging in a north/south direction at noon in Boulder, Colorado at (exactly) latitude = 40^o . Let the Earth rotate about its north pole for a full day back to its initial position with respect to "fixed stars." The plane of the pendulum is seen to rotate through a net angle of 2π -α radians (or 2πcos(40^o) = 276^o).

As an interesting aside, notice that this "deficit angle" α "happens" to be the same as the solid angle of the spherical cap for the given latitude.

That is Ω(lat cap) = ∫ d(area)= ∫ (Rdθ)(2π Rsinθ) from 0 (the north pole) to θ_o and setting R = 1 for a unit sphere. This result is

$\Omega = 2\pi$ (1-cosθ) which was the same as the deficit angle α. Or apply part of the "Gauss-Bonnet theorem" from Stokes Theorem where $\Omega = \int (KdA=dA/R^2 = d\Omega)$. The "K" here is "Gaussian curvature" of the sphere, K = 1/radius^2.

A basis vector e_θ on a northern latitude circle will always point south and be perpendicular to the circle. For any given e_θ at some location on the circle, an adjoining e_θ at angle +dϕ away will seem to precess counter-clockwise.

Then total vector rotation about the cone is $2\pi - \alpha = 2\pi - (2\pi\,(1-\cos\theta))$ = $\mathbf{2\pi\cos\theta}$
Or, rate of rotation is $\omega = \Delta$rotation/Δradians = cosθ .

If we begin with a vector $\mathbf{e_\theta}$ pointing southward, then

$\mathbf{de_\theta/d\phi = \cos\theta}$ or $de_\theta = \cos\theta d\phi = \omega_{\theta\phi}$. **Eqn. 6**

(This $\omega_{\theta\phi}$ is called a connection 1-form and is also derived in Detail A below).

So, this tells how one base vector e_θ rotates, and there is a similar equation for base vector e_ϕ. What it is missing is how much rotation occurs with distance along the phi axis rather than with just angle phi (variable vs directed distance). This is the job of the gradient term with its commensurate distance coordinates. So here, the denominator dϕ → rsinθdϕ = h_ϕ dϕ,and for S^2, we could set r =1). Then we get the result $\nabla_{e\phi}\, e_\theta$ = cosθ /rsinθ = **cot θ/r.**

This agrees with the usual metric index procedure. A primary text on general relativity (MTW [6] p. 213) with orthonormal [ON] unit vectors says,

$\nabla_\phi e_\theta = \Gamma^\phi{}_{\theta\phi}\, e_\phi = \cot\theta\, e_\phi/r$, and $\nabla_\phi e_\phi = \Gamma^\theta{}_{\phi\phi}\, e_\theta = -\cot\theta\, e_\theta/r$ [on S^2]. **Eqn 7**.

These "Christoffel" connection Γ coefficients for a sphere are calculated from the usual tedious (general relativity type) derivatives of the metric tensor [$ds^2 = r^2\, d\theta^2 + (r\sin\theta)^2\, d\phi^2$] – additionally yielding the result : $\Gamma^\theta{}_{\phi\phi}$ = -sinθ cosθ (details not shown here). But, these standard cookbooking index-manipulating calculations aren't intuitively clear, and it would be nice to supplement them with a variety of other simpler approaches.

Once the connection Γ 's are calculated, then the "Riemann tensor" [R_{abcd}] can be formed by derivatives and products of Γ's -- and "Ricci tensor" [R_{ab}] from its "contraction" and Ricci Scalar from one more contraction [R]. All of this is now easily calculated with computer software packages (such as Maple). And all of these are all zero for "flat" space and have simple forms for low dimensions: Riemann is simply combined scalar curvature and metric in 2d and just Ricci and metric in 3d. The Ricci curvature tensor of a 2-manifold surface is just $R_{ab} = Kg_{ab}$ where K is "Gaussian" curvature (which is $1/r^2$ for S^2). These topics are rarely discussed in Euclidean E^3 spaces. But, they are the foundation of 4-d GR geometries.

Once we know how the base vector rotate, we can find the covariant derivative of vectors, $\mathbf{A} = A^i \partial_i = A^i e_i$, so $\partial \mathbf{A}/\partial x^k = e_i \partial A^i/\partial x^k + A^i \partial e_i/\partial x^k = e_i\, \partial A^i/\partial x^k + A^i \Gamma^\gamma{}_{ik} e_\gamma$. [20]. The γ (gamma) and i are dummy indexes to be summed over—their names can be interchanged: Then

$\partial \mathbf{A}/\partial x^k = (\partial A^i/\partial x^k + A^\gamma \Gamma^i{}_{\gamma k})e_i$ or $\nabla_k A^i = (\partial A^i/\partial x^k + A^\gamma \Gamma^i{}_{\gamma k})$ **Eqn. 8**

For vector field in R^3, for "all low" indices (orthonormal and unit frame bases E): if $\mathbf{A}$ = $f_1E_1 + f_2E_2 = f_i E_i$ [1, p 338]:

$\nabla_k A = (E_k\cdot\nabla f_1 + f_2\omega_{21}(E_k))E_1 + (E_k\cdot\nabla f_2 + f_1\omega_{12}(E_k))E_2 = (\partial f_i/\partial x^k + f_j\, \omega_{ji})E_i$

Next (Case 2): We look at the Differential Geometry view [1] showing that the "connection" for this is defined by rotation "forms" ω – but the usual math (see Appendix) is not quite trivial (and the above picture is easier). The rotation of a frame unit vector E_i turning towards another unit vector E_j is given by the "covariant derivative" of the basis vector:

$\nabla_v E_i = \Sigma\ \omega_{ij}(v)\ E_j$ (v is any tangent vector at a point p on a surface).

So $\nabla_v e_\theta = \omega_{\theta\phi}\ e_\phi$ –with the "1-form" $\boldsymbol{\omega = \cos\theta d\phi}$ (e.g., Eqn 11 below). Initially this looks different from the cotθ above. That is, a theta vector pointing down will tilt towards a phi vector to the right (a counter-clockwise rotation). Thinking in "differential forms" is an old but powerful concept that is generally unknown to most students. Now, since the **d**ϕ above is a 1-form basis, it is also an example of $e^\mu = \nabla u_\mu = e_\mu/h_\mu$, so, e.g.,

$\omega_{\theta\phi} = \cos\theta d\phi \rightarrow \cos\theta\ |e_\phi\ /r\sin\theta| =$ **cotθ** $/r = \Gamma^\phi{}_{\theta\phi}\ e_\phi$ .. (so we have consistency).

{there are a variety of notations for ω like $\omega^j{}_i$ (see appendix at end) or $\omega^i{}_{jk}$ -- as in $\nabla_{ej} e_k = e_i \omega^i{}_{jk}$, and if $e_j = \partial_j$, then $\omega^i{}_{jk} = \Gamma^i{}_{jk}$ [11, p 243]. $\omega^k{}_j = \omega^k{}_{rj}\,\sigma^r$ where σ is a 1-form like ***d****x. We can exchange the jk in Γ (symmetry) }.*

Case 3: Another approach to base rotation is to exploit a clever trick of beginning with a constant basis vector [e.g., 3, p 355] to get a turning connection for contravariant bases:

On S^2, e_ϕ (unit) = $\partial_\phi/r\sin\theta$, and $\nabla_{\partial\theta}\, e_\phi = 0$ (no change in unit phi with change in latitude). So, $\nabla_{\partial\theta}\ (\partial_\phi/\sin\theta) = 0 = \nabla_{\partial\theta}\ \partial_\phi\ /\sin\theta + \partial_\phi\ \nabla_{\partial\theta}\ (1/\sin\theta)$. But the last term is $\partial_\theta \cdot (e_\theta \cos\theta/\sin^2\theta)\ \partial_\phi$, so $\nabla_{\partial\theta}\ (\partial_\phi) = \cot\theta\ \partial_\phi$ -- notice this is a different from the $\Gamma^\phi{}_{\theta\phi}$ in Equation 7 because it is now done in a coordinate basis. Unit vectors are not needed to find connections, the non-unit bases work (and are preferred).

Differential Forms:

A 1-form is simply "an expression obtained by adding and multiplying real-valued functions and the differentials dx_1, dx_2…" [1]. So, 3xdx is a form and yzdx + 2dz is a form. They are also considered as the integrand appearing under an integral sign, ∫ . An added (anti-symmetry) rule is that the order of the differentials counts: dxdy = -dydx --- which also implies that dxdx = $(dx)^2 = 0$ (1-form repeats of single objects give zero). As a reminder of this "alternation rule," a wedge symbol may be used: dxdy = $dx \wedge dy$ (called a "Grassman product" or "exterior" product).

If we write a 3d 1-form as $\alpha = a_1dx_1 + a_2dx_2 + a_3dx_3$, it is initially curious that the product of 1-forms for Cartesian coordinates naturally results in a familiar **"cross product"** [11]. $\alpha \wedge \beta \sim (a\cdot dx) \wedge (b\cdot dx) = (a \times b)\cdot dS$

[where, for example, $dS_{12} = dx_1 \wedge dx_2$ is a 2-form]. Unlike the "interior product" (which reduces two vectors to a scalar), the exterior product elevates 1-forms to 2-forms.

This kind of multiplication is similar to that of the "3-vector" part of quaternions [Hamilton's hypercomplex number system] with new "imaginaries" $i^2=j^2=k^2=ijk=-1$; which implies ij=k, jk = I, ki = -j]. Like forms, i j = -j i (antisymmetry). A vector in this basis may look like $u = u_1i+u_2j+u_3k$. If we consistently ignore a usual scalar part addition (q = scalar + vector without the scalar), then the product **uv = (u x v)** .

One more addition to exterior algebra is the concept of **"exterior derivative" d** such that if ψ = fdx + gdy, then we get a 2-form η =**dψ** = df$\wedge$ dx + dg$\wedge$ dy – (and terms like ddx=0). The dx's and dx$\wedge$ dy's are treated as a basis (like e^i), so ψ is similar to a "vector."

$$\mathbf{d} = [(\partial/\partial x)dx + (\partial/\partial y)dy + (\partial/\partial z)dz]\wedge$$

Now things get interesting! It can be shown that **d**f (of a scalar function) is like a gradient f = $\sim\nabla f$, **d**ψ is like a **curl** of vector ψ $\sim\nabla\times\psi$, and **d**η is like a divergence, $\nabla\cdot\eta$. Also **dd**ψ = 0 (just like the divergence of a curl $\nabla\cdot\nabla\times\psi = 0$). And forms can also be integrated. This single symbol **d** serves to include much of vector analysis in easier symbolic form !

If we let vector **A** be the electromagnetic vector potential 1-form in 3-space, then **d**A ~ curl A = $\nabla\times$A= B (the magnetic field).

{Actually a curl vector like B is a "pseudo-vector – it reverses direction under mirror reflection. We should represent "B"=dA as a 2-form (Greek)$\beta = b_{ij}dx^i\wedge dx^j$. But we can then apply a complimentary "3-space Hodge star" operation to look like a 1-form vector: $*\beta = \beta_i dx^i$. Similarly $*(dx\wedge dy)=dz$.
For the space of p-forms labeled as wedge Λ^p, we have $*\Lambda^p=\Lambda^{n-p}$, so $*\Lambda^2=\Lambda^{(3-2=1)}$ or just 1-forms }.

Carrying this over to relativity in 4-space can then give a **"generalized curl"** of a 4-vector, like F = **d**A , where A is now the vector potential 4-vector, A_μ, and **F** is like the relativistic electro-magnetic tensor $F_{\mu\nu}$. One might think that if we can write F=dA so simply then the mechanism of Nature may operate with the same simplicity or something isomorphic to it. The presence of E and B fields in the resulting F is said to represent "curvature" for F. The simple expression **d**F = **dd**A = 0 contains the Faraday law and no-magnetic-poles law of Maxwell equations. But this is also an example of a "Bianchi" identity. That is a concept from general relativity that also applies to EM. With forms, a lot of math and physics can be summarized more compactly than it used to be. One can switch between conventional math and forms using the various powers of forms as needed [e.g., 2-forms can be concretely represented by matrices like the $F\rightarrow F_{\mu\nu}$.]

Invariance is a major theme in differential geometry. Various choices of coordinate systems don't matter – none is intrinsically preferred. The laws of physics are unchanged by a change in coordinate system, rotations and boosts in speed (called "general covariance"). However, practical physicists initially define almost everything in terms of coordinates and only later show independence. Vector components are not coordinate independent, but the combination $A^\mu\partial_\mu$ is. Mathematicians prefer to define forms without mentioning coordinates and sometimes even without metric. In Euclidean space we have isometries or mappings that preserve the distance between two points regardless of rotation and translation (like congruences between two triangles in plane-geometry). In special relativity, we have the electromagnetic tensor F that can show a preservation of Maxwell equations (for Lorentz invariance). F is anti-symmetric and is also a "2-form" resulting from a generalized curl of a vector potential (F = **d**A). The (Lorentz) "invariants" of F are : $(F^{\mu\nu}F_{\mu\nu}) = E^2 - B^2$ and the dot product $E\cdot B$.

Tensors in general exist independently of any frame of reference or choice of coordinates and are defined by their transformation laws. A vector is a first-rank tensor and has only one invariant, its length. The symmetric metric tensor, $g_{\mu\nu}$, is of second rank and expresses the geometry of space (or space-time). It is the inner product of two tangent vector bases. For us, for the surface of S^2, we have:

$g_{\mu\nu} = \langle \partial_{x\mu}, \partial_{x\nu} \rangle = h_\mu e_\mu \cdot h_\nu e_\nu = (h_\mu)^2 \delta_{\mu\nu}$, so

$d\ell^2 = g_{\mu\nu} dx^\mu dx^\nu = R^2 d\theta^2 + (R\sin\theta)^2 d\phi^2$. **Eqn. 9**.

Once we've acquired a feeling for the covariant derivative, it might be mentioned that it is applied twice on a vector , w, to get Riemannian curvature. In the simple case where two basis vectors are defined as $u = \partial / \partial x^i$ and $v = \partial / \partial x^j$,
Riemann $R(u,v)\, w = [\nabla_u, \nabla_v] w = \nabla_u \nabla_v w - \nabla_v \nabla_u w$. This "commutator" measures the "noncommutativity of the covariant derivative" [2].

Other Examples of Covariant Exterior derivative: 1-form, 2-forms, 3-forms.

The symbol "**D**" for covariant derivative is used to include both derivatives and connections (Γ's or ω's).
For relativistic electricity and magnetism fields (4-vector EM), the covariant derivative is: $\mathbf{D}^\mu = \partial^\mu + iqA^\mu(x)$ for the group U(1) with "A" or "qA" being referred to as a connection. One author calls this connection a "twist in Minkowski space" [24]. The vector potential A field can alter the phase of a moving electron. [**I** think that qA represents "the dragging of electromagnetic 'space' by distant moving, circulating or spinning charges."]
In terms of the U(1) group, this means that an action term like $\int_\gamma A$ (integrated over the path gamma) becomes $\exp(i \int_\gamma A) \in U(1)$ – a complex phase term.

With more detail, E and H are technically 1-forms, D, B, and J are 2-forms (e.g., $J_1 dy \wedge dx$), and charge $Q = \rho\, dxdydz$ is a 3-form. Then Maxwell's equations are:

$\mathbf{d}E = -\partial B/dt$, $\mathbf{d}H = \partial D/\partial t + J$, $\mathbf{d}D = Q$, and $\mathbf{d}B = \mathbf{dd}(\text{vector } A) = 0$.

For the Weak interactions, this becomes something much more complex. This is the sense in which A is a U(1) connection. The Faraday tensor F=dA says that the EM field is the curvature of the A-connection. F is used in equations like $m(\partial u^j/\partial t) = qF^j{}_k u^k$ bringing together charge (q), mass (m), and velocity into a form of the "Lorentz force law" (separate from the usual Maxwell equations). Physicists and mathematicians see connections differently: "physicists don't know why their notion of a field should play a role in the description of parallel transport, and the mathematicians don't know why the object they use to describe parallel transport should have any physical significance as a field [24]."

For weak interactions, we generalize the covariant derivative from electromagnetism to $D^\mu = \partial^\mu + ig\, \boldsymbol{\tau} \cdot W^\mu(x)/2$ using the group SU(2) where tau is like Pauli matrices $\sigma^a/2$. {Actually, in the standard model we go to a combined "Electro-Weak" with group SU(2)xU(1) }. The forms idea here is that the F = dA gets generalized to $F = dA + A \wedge A$

with a more elaborate "A." This is analogous to the general space curvature (2-form) $\Omega = D\omega = d\omega + \omega \wedge \omega$ [see Details Part C below].

This use of weak SU(2) portion goes by the name **Yang-Mills (YM),** and has been a guiding principle of particle physics since 1954. The realization that the weak field W^μ represents a connection didn't occur to Yang until 1975 [24]. Generalizing even this to the group unbroken group SU(3) gives the quantum chromodynamics (QCD) theory for strong interactions. "The standard model [SM] using the combined compact group ~ U(1)× SU(2)×SU(3) is one of the greatest achievements of science and is a quantized Yang-Mills theory [24]."

For a pair of tangent vectors, X, and Y, $R(X,Y) = \Omega(X,Y) = d\omega(X,Y) + [\omega(X), \omega(Y)]/2$, where R is a "curvature tensor" and Ω is a curvature form. In MTW [6] this is written as $\Omega^i_j = d\omega^i_j + \Sigma\, \omega^i_k \wedge \omega^k_j = \mathcal{R}^i_j$ (Script R in MTW)! A connection is called "flat" if Ω =0.

The **Bianchi Identities** of general relativity are a generalization of d(dω) = 0, or $\partial\,\partial\,\omega = 0$ (the boundary of a boundary is zero). **3-forms** may derive from the exterior derivative of 2-forms. Of particular interest here is the "Torsion tensor" 2-form -- a measure of twisting or screwing of a moving frame around a curve or the twisting of tangent spaces around a curve [Eqn 10]. The simplest example of torsion is the twisting of a helix or helicoid such as the "Archimedes Screw." For connections, torsion can be thought of as the lack of symmetry of the Christoffel symbols: $T^k_{i\,j} = \Gamma^k_{i\,j} - \Gamma^{\,k}_{j\,i}$. In General Relativity we almost always assume symmetric connections – no torsion). Instead of the usual bases, e_i, we now consider the dual basis θ^i such that $\theta^i(e_j) = \delta^i_j = 1$ or 0 [see Details B] . As a 2-form, torsion is

$$\Theta^{\,k} = d\theta^k + \omega^k_j \wedge \theta^j = T^k_{ij}\, \theta^i \wedge \theta^j \text{, or } \boldsymbol{\Theta = D\theta}. \qquad \text{Eqn. 10}$$

The two types of Bianchi identities are then: 1. $\mathbf{D\Theta = \Omega \wedge \theta}$, and 2. $\mathbf{D\Omega = 0}$ -- as 3-forms.

The first Bianchi identity in component form for the Riemann tensor is:
$R_{abcd} + R_{acdb} + R_{adbc} = 0$.

A component form for the 2nd Bianchi Identity of the Riemann tensor is $\{R_{\alpha\,\eta\,\beta\,\gamma;\,\delta}\}_{(\beta,\gamma,\delta)} = 0$ [keeping the first two subscripts constant and cycling through the ( ) terms].
Or, $R_{\alpha\,\eta\,\beta\,\gamma;\,\delta} + R_{\alpha\,\eta\,\gamma\delta;\,\beta} + R_{\alpha\,\eta\,\delta\,\beta;\,\gamma} = 0$.

These express symmetries and redundancies of the Riemann tensor (not all of it is useful, so the Einstein equations may use the contracted Ricci tensor.

Details:

A. Connection ω 1-forms: for the selected case of S^2 spherical polar coordinates with unit vectors for r, θ , and ϕ (Frame $E_1, E_2, E_3 = E_r, E_\theta, E_\phi$). Obtain turning coefficient ω forms using the "natural" orthonormal Euclidean frame basis, i, j , k.

$E_1 = e_r$ (with "hat" ^) = sinθ cosϕ i + sinθ sinϕ j + cosθ k = a_{11} i + a_{22} j + a_{33} k.
These coefficients of ijk (or U_1, U_2, U_3) are part of what is called an "altitude matrix" $A = a_{ij}$ where $E_i = \Sigma\, a_{ij} U_j$.
$E_2 = e_\theta$ = cosθ cosϕ i +cosθ sinϕ j –sinθ k
$E_3 = e_\phi$ = -sinϕ i + cosϕ j. Then differentiate all of these and consolidate terms to get:

de_r = i(cosθ cosϕdθ –sinθ sinϕdϕ) + j(cosθ dθ sinϕ +sinθ cosϕ dϕ) –ksinθdθ
=dθ (e_θ) + dϕ (sinθ e_ϕ) [$=\omega_{12}e_2+\omega_{13}e_3$]. {"all low" convention}.
de_θ = i(-sinθ cosϕ dθ –cosθ sinϕ dϕ)+j(-sinθ sinϕ dθ +cosθ cosϕ dϕ)-kcosθdθ
= -e_r dθ +cosθ e_ϕ dϕ [$=\omega_{21}e_1+\omega_{23}e_3$]
de_ϕ = -icosϕdϕ –sinϕ j dϕ =-sinθ e_r dϕ –cosθ e_θ dϕ. {using $\sin^2\theta + \cos^2\theta = 1$}.
[$=\omega_{31}e_1+\omega_{32}e_2$] . **Eqn. 11.**

With these coefficient assignments for omega's, we can now read off the connection forms for S^2: $\omega_{12} = \omega_{r\theta}$ = dθ , $\omega_{13} = \omega_{r\phi}$ = sinθdϕ , $\omega_{23} = \omega_{\theta\phi}$ = cosθdϕ.
And as anti-symmetric forms we have $\omega_{21}=-\omega_{12}$, $\omega_{13} = -\omega_{31}$, and $\omega_{32} = -\omega_{23}$ (order counts! e.g., base 1 turning towards base 2 implies that base 2 turns away from base 1). Here, we have used "orthonormal" frames. If instead $e_j = \partial_j$ (not unit lengths), then $\omega^i_{jk} = \Gamma^i_{jk}$ with different results inter-related by h_i's . If spherical radius is a constant, then $de_\theta = \omega_{\theta\phi}\, e_\phi$.

{ Note: The process above is equivalent to finding an antisymmetric connection matrix $\omega = dA\, A^T$ from the "altitude" matrix of unit frame coefficients [1].}

A key equation in differential geometry is: $\nabla_v E_i \sim v\cdot \nabla E_i = \Sigma\, \omega_{ij}(v) E_j$ for all tangent vectors v. The implication from this is that our de_i's must be equivalent to the $\nabla_v E_i$'s. (with the old confusion of thinking of d as "tiny" Δ ' s). Indeed, for the previous latitude cone construction, it was true that $de_\theta = \omega_{23}$ (e_ϕ): i.e. **de_θ/dϕ = cosθ.** where $\omega_{23} = \omega_{\theta\phi}$ = cosθdϕ. In this view we indeed have tiny d expressing an increment of basis rotation. **d**f is often used for expressing ∇f (e.g., $\partial_v f = \langle df, v\rangle = \nabla f\cdot v$ [MTW p60]).
Also, in terms of Christoffel symbols:

$\omega^k{}_i \equiv \Gamma^k{}_{ij}\, \sigma^j$ (where σ can be just another symbol for our previous "θ" 1-form bases).
For our sphere example,
$\omega^\phi{}_\theta = (\cot\theta / r)(\theta^\phi = r\sin\theta d\phi) = \cos\theta\, d\phi$ (as before above).

B. Dual Forms:

From O'Neil [1], the dual 1-forms for sphere S^2 are **$\theta_i = h_i\, dx_i$**: so θ_1=dr, θ_2=rdθ , θ_3 = rsinθ dϕ (the h_i's are the scale factors—and the θ's are like our older "unit" vectors e's – and MTW would call them ω^i with little hats on them for unit lengths). O'Neil uses orthonormal bases, so his duals are with-respect-to unit vectors and $\theta_i=\theta^i$. The scale factors serve to make the forms commensurate as distances. The metric tensor can be expressed using duals. For the case of the unit sphere: $g = d\theta^2+\sin^2\theta d\phi^2 = (\theta^\theta)^2+(\theta^\phi)^2$.

Some books introduce base transformations called 2-index "veirbeins" (for 4d space) to transform from coordinate to "non-coordinate" (orthonormal) bases: $e_i = e_i^{\mu} \partial_{\mu}$. They may also transform dual bases: $\theta^i = e^i_{\mu} dx^{\mu}$ (and $e^i_{\mu} e_i^{\nu} = \delta_{\mu}^{\nu}$ & $e^i_{\mu} e_j^{\mu} = \delta^i_j$ [23]).

For the simple case of the sphere S^2, $\theta^{\theta} = e^{\theta}{}_i dx^i = e^{\theta}{}_{\theta}\, d\theta + e^{\theta}{}_{\phi} d\phi = r d\theta$ or just $d\theta$ for r = 1 unit sphere. Also $\theta^{\phi} = e^{\phi}{}_i dx^i = e^{\phi}{}_{\theta}\, d\theta + e^{\phi}{}_{\phi} d\phi = r\sin\theta d\phi$. Then one can just read off two non-zero "zwei-beins" as $e^{\theta}{}_{\theta} = 1$ and $e^{\phi}{}_{\phi} = \sin\theta$. These are related to what are called "tetrads" in general relativity. Notice that these transformations accomplish the same thing as our previous h_i's.

Names of terms can be tricky [21] – meaning "new" for students of vector analysis: The understood standard **"coordinate frame"** for S^2 is $e_i = \partial / \partial u^i$ (e.g., $e_{\phi} = \partial / \partial \phi$), with "dual frame" [following MTW [6]] $\boldsymbol{\omega}^i = du^i$ (e.g., $\omega^{\phi} = \mathbf{d}\phi$ – note that this gradient like basis compensates for the tangent like basis $e_{\phi} = \partial / \partial \phi$ – and neither has unit length.) These obey the dual-rule $du^i(e^i) = \omega^i(\partial_i) = \delta^i_j$= 1's or 0's.

Then there are the **"associated orthonormal frames"** like $e_{\phi}(\hat{}) = (\partial / \partial \phi)/h_{\phi} = (\partial / \partial \phi)/r\sin\theta$ (with index phi-hat = ϕ ^). And the orthonormal duals are $\theta^{\phi}(\hat{}) = r\sin\, d\phi = h_i\, du^i$. Both of these pairs obey $\theta(e) = \delta^i{}_j$, and with e having unit length.

That is, dual 1-forms are defined so that $\boldsymbol{\theta_i(v) = v\cdot E_i(p)} = v_i$ [1 p 94 for orthonormal Frame bases]. So, as in linear algebra, the θ's are linear functionals, and here they project a tangent vector onto its i'th component to yield a scalar length. If vector v is chosen to be E_j, then $\boldsymbol{\theta_i(E_j)=\delta_{ij}}$. This is also a general requirement for dual frame bases, and the θ_i's above are the dual basis of E_i. { example: $\theta_3(E_3) = 1 = r\sin\theta d\phi(e_{\phi})$ is a projection onto the ϕ unit tangent vector. BUT, that unit tangent **should be in terms of coordinate frame** $\partial / \partial u^i$,

SO, $\theta_3(E_3) = \theta^3(E_3) = \theta^{\phi}(E_{\phi})\ (r\sin\theta d\phi\ )(\partial / \partial \phi / r\sin\, d\phi) = d\phi\ (\partial / \partial \phi) = 1$.

The gradient [eqn. 5] should really be written using these dual bases:
$\nabla f = \mathbf{d}f = \Sigma\ (\partial f / \partial u^i)\omega^i = \Sigma(\partial f / \partial u^i)\ (\theta^i/h_i)$. *[MTW p 206].*
{instead of the older unit vectors e^}.
Directional derivatives and covariant derivatives use gradients and therefore should be using dual bases. When we say connection $\omega_{\theta\phi} = \cos\theta d\phi$, the $d\phi$ = unit $\omega^{\phi}/r\sin\theta$ = θ^i/h_i *.*
*Then, $\omega_{\theta\phi} = \cos\theta d\phi = \cos\theta\ \theta^{\phi}/r\sin\theta = \theta^{\phi}$ **$\cot\theta/r$** $= \Gamma^{\phi}{}_{\theta\phi}\, \theta^{\phi}$ for orthonormal bases. !!*

The difference is the use of the dual basis for 1-forms.
{And a reminder that in coordinate basis $\Gamma^{\phi}{}_{\theta\phi} = \cot\theta$ (without the / r). So, $\nabla_{\partial\theta}(\partial_{\phi}) = \cot\theta\, \partial_{\phi}$.

For the natural frame field, ijk, the operation $dx_i(v) = v_i = v\cdot U_i(p)$ implies that θ_i for that Euclidean basis is $\mathbf{d}x_i$ (the simplest 1-forms). Then, any tangent vector can be written as $v = v_i E_i = \Sigma\theta_i(v)E_i$. 1-forms themselves may also be expressed using basis forms: $\phi = \Sigma\ \phi(E_i)\ \theta_i$. {or as linear functionals, $(\Sigma\ \phi(E_i)\ \theta_i)(v) = \phi(v)$ }.

OK, so why is any of this important? It eventually enables us to calculate the Gaussian curvature (K) of a surface in terms of local behavior of ω_{ij} and θ_i's.

For Euclidean 3-space, we work with 2d surfaces. We prefer to use an **"adaptive frame"** in which the first two bases (1 and 2) are on the surface and the third is normal to the surface, $\perp$. For the sphere, the order r,θ,ϕ is cyclically shifted to 1,2,3 = θ ,ϕ ,r with only

$\theta_1 = \theta_\theta = r d\theta$ and $\theta_2 = \theta_\phi = r\sin\theta d\phi$ forms being used.

The turning connections are then $\omega_{12} = \cos\theta d\phi$, $\omega_{\theta r} = \omega_{13} = d\theta$, and $\omega_{23} = -\sin\theta d\phi = \omega_{\phi r}$.

There is a major concept called "Riemann curvature" expressed with four indices: R_{abcd} (or $R^i_{jk\ell}$). It is very important in 4d general relativity space-time. It can measure the "extent to which the metric tensor is not locally isometric to that of Euclidean space" and the "noncommutativity of the covariant derivative" and also "the non-holonomy of the" Riemannian manifold.

But in a 3-d space, it can be expressed solely in terms of something smaller: the 2-index "Ricci" tensor R_{ab} along with the metric tensor g_{ab}. In 2-d space, it can be expressed in terms of just the contracted Ricci scalar $R = R^a_a$ (no index left, [16]). Actual calculation for S^2 gives Riemann $R_{\theta\phi\,\theta\phi} = r^2\sin^2\theta = g_{\theta\theta} R^\theta_{\ \phi\,\theta\phi}$, so $R^\theta_{\ \phi\,\theta\phi} = \sin^2\theta$.

This form of the Riemann tensor has direct relationship to similar expressions from abstract differential forms:

C. The "Curvature 2-form" Ω is:

$\Omega^i_j = d\omega^i_j + \omega^i_m \wedge \omega^m_j = (1/2) R^i_{jkl}\sigma^k \wedge \sigma^\ell$, or $\Omega^i_j(X,Y)e_i = (1/2)\,Rm(X,Y)e_j$. **Eqn. 12.**

Or in simpler abstract notation: $\boldsymbol{\Omega = D\omega} = \boldsymbol{d\omega + \omega \wedge \omega = d\omega + \omega^2}$.

[different sources may vary on whether a "-" sign of + sign is used]. Wald's notation for curvature 2-form $\boldsymbol{R}_\mu^{\ \nu}$ [eqn. 1] is the same as our Ω.

For the simple 2d spherical surface, we either have $\omega^2 = 0 \wedge \omega$ or $\omega \wedge 0 = 0$. This is due to having only two variables θ and ϕ to play with and $\omega^\theta_\theta = \omega^\phi_\phi = 0$ [remember that both ω and Ω are antisymmetric – no diagonal entries].

So, for $\mathbf{S^2}$ we are left with simply: $\Omega^i_j = d\omega^i_j = \sin\theta\, d\theta \wedge d\phi$. That is, the curvature form for S^2 only has two off diagonal entries: Ω^i_j and $\Omega^j_i = -\Omega^i_j = -\sin\theta d\theta \wedge d\phi$.

This form [eqn. 12] may be derived by taking **two consecutive exterior derivatives** of a vector: **ddv =dd**$(e_\mu v^\mu) \rightarrow$ $\mathbf{d^2 v} = e_\mu \Omega^\mu_{\ \nu} v^\nu$ which is linear in v [6, MTW, derivation not shown here].

A simpler approach is to just take $\mathbf{d^2 (e_j)}$ for an orthonormal unit vector.

$d^2e_j = d(e_k\omega^k_j) = de_k \wedge \omega^k_j + e_k d\omega^i_j = e_i(\omega^i_k \wedge \omega^k_j + d\omega^i_j) = e_i\Omega^i_j$.

Or, more abstractly: $\nabla\nabla e = \nabla(e\omega) = (\nabla e)\omega + e d\omega = e(\omega \wedge \omega + d\omega) = e\Omega$ [11].

Now, Ω is stated in books to be $= R^i_{jkl}\sigma^k \wedge \sigma^\ell$ and $R^i_{jkl} = \sin^2\theta$ for coordinate bases. But ω's were defined for orthonormal bases (unit vectors).

[Many sources agree that $\Omega^i_j = d\omega^i_j = \sin\theta\, d\theta \wedge d\phi$ for S^2, and we may also see this from duals: $d\omega_{12} = -K\theta_1 \wedge \theta_2$ [1 p270] . That is:

$-\sin\theta d\theta\, d\phi + \cos\theta\, (dd\phi=0) = K[\, r d\theta \wedge r\sin\theta\, d\phi = r^2 \sin\theta\, d\theta \wedge d\phi\,]$
So – $K=-1/r^2$ or $\mathbf{K = 1/r^2}$. {the usual Gaussian curvature}.

Many sources say that $R^{\theta}{}_{\phi\theta\phi} = \sin^2\theta$ without stating their convention. For **orthonormal** bases the tensor $R^{\theta}{}_{\phi\theta\phi} = 1/r_o^2 = K$ – a quite different result. Then,
$\Omega^i{}_j = R^i{}_{jkl}\, \sigma^k \wedge \sigma^\ell$, or $\Omega^{\theta}{}_{\phi} = (1/r_o^2)(\theta^\theta = r_o d\theta)(\theta^\phi = r_o \sin\theta d\phi) = \sin\theta d\theta \wedge d\phi = d\omega^{\theta}{}_{\phi}$. Eqn. 12.
It Works for the orthonormal convention.

For "**Ricci**" (coordinate bases) $R_{ij} = R^k{}_{ikj}$ is diagonal with S^2 values of:

$R_{\theta\theta} = 1$ and $R_{\phi\phi} = \sin^2\theta$, and the **scalar R** $= g_{ij}R^{ij} = 2/r^2 = 2K$ [17].

{Riemann over two tangent vectors X and Y together and for orthonormal bases e_i}.
But our Euclidean space structural equation was $d\omega_{ij} = \Sigma\, \omega_{ik} \wedge \omega_{kj}$ which means that $\Omega^i{}_j = 0$. Why? – because we are studying an introductory low dimensional space which lacks this type of interesting 2-form curvature.

$$V_{\alpha;\beta\gamma} - V_{\alpha;\gamma\beta} = R^{\rho}{}_{\alpha\beta\gamma} V_\rho,$$

corresponds to the following:

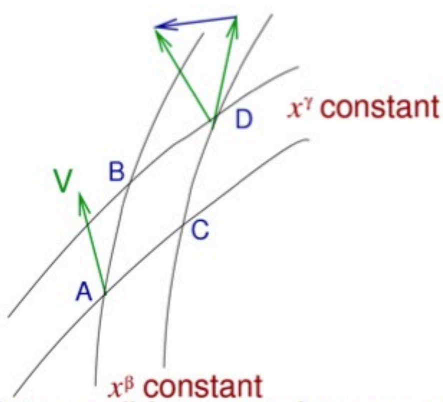


Figure: Vector parallel transported two ways around the same loop does not match up at the end if there is curvature

Vector $\vec{V}$ is first parallel transported $A \to C \to D$, associated with $V^{\alpha}{}_{;\beta\gamma}$. Then the same vector is taken $A \to B \to D$, associated with $V^{\alpha}{}_{;\gamma\beta}$. Curvature causes the vectors at D to differ.

Figure 1. Riemann [Ref 19].

"The vanishing of the Riemann tensor is both a necessary and sufficient condition for Euclidean flat space."

The Three-Sphere, S^3 (HyperSphere) :

So, let's increase the dimension to S^3 embedded in R^4.
The 3-sphere in Cartesian coordinates is formed from $x_1^2+x_2^2+x_3^2 + x_4^2 = R^2$.

Or for angular coordinates, we have three angles on a unit sphere,
(say ψ ,θ ,ϕ with R = 1) with metric:

$d\ell^2 = R^2 (d\psi^2 + \sin^2\psi d\theta^2 + \sin^2\psi \sin^2\theta d\phi^2)$ = diagonal $g_{\mu\mu}\, dx^\mu dx^\mu$. The scale factor for $h_\phi = \sin\psi \sin\theta$ and $h_\theta = \sin\psi$.
After grinding through all the Γ connection derivatives and contraction (not shown here) , we get a **Ricci** Tensor $R_{\mu\nu}$ which is also diagonal: $R_{\psi\psi} = 2$, $R_{\theta\theta} = 2\sin^2\psi$, $R_{\phi\phi} = 2\sin^2\psi \sin^2\theta = 2\, g_{\mu\nu}$ $\{/R^2\}$

The Ricci Tensor is proportional to the metric tensor.
The Ricci Scalar is R = 2+2+2 = 6.
The non-zero elements of the **Riemann** Tensor are also the coefficients of the metric tensor. $R^\Psi\theta_{\psi\theta} = \sin^2\psi = g_{\theta\theta} = (1/2)R_{\theta\theta}$, $R^\Psi{}_{\phi\psi\phi} = (1/2)R_{\phi\phi} = R^\theta{}_{\phi\theta\phi}$.

The 3-form volume is $dV = R^2(\sin^2\psi \sin\theta)d\psi \wedge d\theta \wedge d\phi$.

The Electromagnetic Vector Potential A:

A is a “4-vector,” A = (ϕ,A). As a 1-form: $A = A_\alpha(t,\mathbf{x})\, dx^\alpha$. And $F_{\alpha\beta} = A_{\beta,\alpha} - A_{\alpha,\beta}$
[We assume here the MTW [6] g-metric signature (- +++) --(largely for more positive terms – although my personal preference is generally for $g_{oo} > 0$]. Then the metric - +++ for g means that $A_o = g_{oo}A^o = -A^o = -\phi$ (!) As a 1-form, $A = A_o dt + A_1 dx + A_2 dy + A_3 dz$.
The 4x4 antisymmetric **electromagnetic tensor F 2-form** can be produced simply by one exterior derivative: “**<u>Faraday</u>**” = **F = dA** = **-d**ϕ dt+ $\mathbf{d}A_1$ dx + $\mathbf{d}A_2$ dy + $\mathbf{d}A_3$ dz because terms like dxdx = 0. We may also write out terms as:

$$\mathbf{dA} = (\partial_\mu A_\nu)\, dx^\mu \wedge dx^\nu = (\partial_\mu dx^\mu) \wedge (A_\nu dx^\nu)$$

Now the coefficients A_α are each scalar functions of t,x,y, and z so that dA_α is a 4- term gradient, and the result is a lot of terms. Writing all that out and grouping similar terms together yields terms like $(\partial A_2/ \partial x - \partial A_1/ \partial y)\, dx \wedge dy$ – a 2-form that can be identified as a β_3 part of $B = \nabla \times A$. In a 4x4 matrix this can be placed in row 1 (for dx) and column 2 (for dy) The electric vector field comes from $E = -\nabla \phi - \partial A/ \partial t$, so terms like $(\partial \phi/ \partial x + \partial A_1/ \partial t)\, dt \wedge dx$ stand for $-E_1$ which can be placed in the 4x4 matrix at row 0 (for dt) and column 1 (for dx).

This agrees with the signs for $F_{\alpha\beta}$ in the preferred MTW text and Frankel texts (top row has minus E field).

So, using the B or β 2-form, $F = B + E \wedge dt$ and $dF = dB + dE \wedge dt = ddA = 0$.

{Wikipedia (for the Aharonov-Bohm Effect) says that curvature is $iF = \nabla \wedge \nabla$ where the U(1) connection is $\nabla = \mathbf{d} + iA$. But the AB effect is seen exterior to a solenoid where B=0 and F = 0 (there is no “curvature” there). The connection is similar to the QED covariant derivative term $(-i\hbar\nabla - e\vec{A})\Psi$. }

<u>D. Cartan Structural Equations</u> <u>with</u> subscripts and superscripts [Ref 18] :

For curved space orthonormal frames $\{E_i\}$ with dual basis forms $\{\theta^j\}$ and connections ω^j_k, we again have the incremental frame rotations
$dE_k = \omega^j_k E_j$ (by definition of ω -- notice how the sub and superscripts are now balanced out).
And, **dx** = $d(x_1, x_2, \ldots x_n) = \theta^j E_j$ with dual x_j functions defined as $x_j(p)=p_j$.

Now Differentiate these equations:
$d(dx) = d^2x=0 = d(\theta^j E_j) = d(E_j\theta^j) = E_j d\theta^j + dE_j \wedge \theta^j = E_j d\theta^j - \theta^j \wedge dE_j =$
$d\theta^j E_j - \theta^\ell \wedge \omega^j_\ell E_j = (d\theta^j - \theta^\ell \wedge \omega^j_\ell) E_j = 0, \Rightarrow d\theta^j - \theta^\ell \wedge \omega^j_\ell = 0.$
So, $d\theta^j = \theta^\ell \wedge \omega^j_\ell$. The First Structural Equation. Eqn. 13a.

[in all these cases, the ω upper and lower indices are supposed to be vertically aligned rather than shifted sideways from each other – but Word can't do that]
Similarly, $d(dE_k) = 0 = d\omega^j_k E_j - \omega^j_k \wedge dE_j = d\omega^j_k E_j - \omega_k^\ell \wedge \omega^j_\ell \wedge E_j =$
$(d\omega^j_k - \omega_k^\ell \wedge \omega^j_\ell) E_j = 0$.
So, $d\omega^j_k = \omega_k^\ell \wedge \omega^j_\ell$. The Second Structural Equation. Eqn. 13b.

The "all low" orthonormal bases convention from O'Neil had: $\mathbf{d\theta_i = \Sigma\, \omega_{ij} \wedge \theta_j}$. {e.g., $d\theta_3 = d(\sin\theta\, d\phi) = -\cos\theta d\theta\, d\phi = \omega_{32} \wedge \theta_2 = (-\cos\theta\, d\phi) \wedge d\theta$ } ???
Also, $d\omega_{ij} = \Sigma\, \omega_{ik} \wedge \omega_{kj}$. {for the "**second** structural equation"}. This is also written as: $d\omega^i_j = \omega^i_k \wedge \omega^k_j$ (orthonormal frames).

Now consider an R^3 "adapted frame" where E_1, E_2 and θ^1, θ^2 are required to be **on** the surface (symbol ⊤) but E_3 is now normal (⊥) to it. For the sphere case, S^2, we simply cyclically rotate the trio (r, θ ,ϕ) → (θ ,ϕ, r) =(i=1,2,3) where the angles are on the surface and r is naturally perpendicular to it. The third dual is now gone, $\theta^3 = 0$ because it is a linear functional that operates on tangent vectors, v, which are parallel to the surface: $\theta^3(v) = v\cdot E_3 = 0$.

$dE_i=\omega_i^j E_j + \omega_i^k E_k$ (with i,j,k cyclic). So, e.g., $dE_1=\omega_1^2 E_2 + \omega_1^3 E_3$. And for the duals, $d\theta^1 = \theta^2 \wedge \omega^1_2$, and $d\theta^2 = \theta^1 \wedge \omega^2_1$.
And $\theta^3=0 \Rightarrow d\theta^3 = 0 = \theta^1 \wedge \omega^3_1 + \theta^2 \wedge \omega^3_2$
Re-applying the 2nd structural equations in the adapted frame gives special cases with new names called the: **"Gauss Equation"** [18] : $d\omega_1^2 = \omega_1^3 \wedge \omega_3^2$ Eqn.14a, b.
and two "Codazzi equations:" $d\omega_1^3 = \omega_1^2 \wedge \omega_2^3$ and $d\omega_2^3 = \omega_2^1 \wedge \omega_1^3$

Finally Curvature:

An introduction to **Gaussian curvature** begins with observing how much a normal unit vector to a surface "falls" forward with movement on the surface. This is extrinsically described (by something called the "shape operator, S."
The **Shape operator** $S_p(v) = -\nabla_v U$, where U is a **unit normal** to M [1]. This is extrinsic geometry since a normal lies beyond the dimensions of a surface (embedding). For an "adapted" frame, the third direction E_3 **is** that unit normal vector: $S_p(v) = -\nabla_v E_3$.
For our sphere, U = unit r = vector r/|r| = $(x_i U_i)/r$. Then $\nabla U = 1/r$ and $S_p(v) = -v/r$. So S is simply multiplication by -1/r = -k.
If **u**= unit tangent, then

$S_p(\mathbf{u})\cdot \mathbf{u} = -1/r \equiv k(\mathbf{u})$ for a **"normal curvature"** with respect to the **u** direction on the surface. For S^2, all tangent directions give the same curvature.

In general, for principle directions e_1 and e_2 on a surface, $S(e_1) = k_1e_1$, $S(e_2) = k_2e_2$ may be expressed as a diagonal matrix $S_{11} = k_1$ and $S_{22} = k_2$. Then Gaussian curvature is the determinant of S: $K = \det S = k_1k_2$.

Since $\nabla_v E_3$ describes a frame base rotation along v, shape is also described by ω turning forms. $\mathbf{S= \omega_{13}E_1 + \omega_{23}E_2;}$ Eqn. 15a

and all the Cartan dual forms and connections apply.

Then we can show [1, p270] that: $\mathbf{K\theta_1\wedge\theta_2 = \omega_{13}\wedge\omega_{23} = -d\omega_{12}}$. **Eqn. 15b**

(kind of a 2^{nd} derivative of E_1, E_2).

That is: $S(E_1) = -\nabla_{E1}E_3 = -\omega_{31}(E_1)E_1 - \omega_{32}(E_1)E_2$

and $S(E_2) = -\nabla_{E2}E_3 = -\omega_{31}(E_2)E_1 - \omega_{32}(E_2)E_2$

The determinant of the 2x2 matrix of these ω terms is

$\det S = \omega_{13}(E_1)\omega_{23}(E_2) - \omega_{13}(E_2)\omega_{23}(E_1) \equiv K$.

But the 2-form on 2 frame bases $(\omega_{13}\wedge\omega_{23})(E_1,E_2)$ which also equals this determinant expansion. And this also gives the Gauss equation for $d\omega_{12}$.

And, as a 2-form, the duals obey $(\theta_1\wedge\theta_2)(E_1,E_2) = \theta_1(E_1)\theta_2(E_2) + 0 = 1\cdot 1 = 1$.

In 3d, for 2d surfaces, $\theta^3 = 0$ and the Curvature 2-form $\mathbf{\Omega = D\omega = d\omega + \omega^2}$ will only depend on the first term, $\Omega^1{}_2 = d\omega^1{}_2 = K\theta^1\wedge\theta^2$.

And for a sphere S^2: $(1/r^2)(rd\theta \wedge r\sin\theta d\phi) =$ $\mathbf{\sin\theta d\theta\wedge d\phi}$ $= \Omega^1{}_2 = d\omega^1{}_2$.

Also, $\Omega^2{}_1 = d\omega^2{}_1 = -d\omega^1{}_2$. So Ω **is a 2x2 skew symmetric matrix of 2-forms** $[^{o}_{-}{}^{+}_{o}]$

[The diagonal $\Omega^1{}_1 = \Omega^2{}_2 = 0$]. The curvature of S^2 is a positive constant: K = 1 [18].

The 2x2 **Ricci** curvature tensor is a somewhat different beast.

$R_{\theta\theta} = 1$ and $R_{\phi\phi} = \sin^2\theta$, and the scalar $R = g_{ij}R^{ij} = 2/r^2 = 2K$. Ricci is symmetric and has entries on the diagonal, while Ω is anti-symmetric with zeros on the diagonal and non-zero entries off of the diagonal.

References:

1. **Barrett O'Neil**, **Elementary Differential Geometry,** 2^{nd}. Ed., Academic Press, 1966, 1997, 2006.
2. Wikepedia on "Covariant Derivative" and "Riemann curvature tensor."
3. William L. Burke, Applied Differential Geometry, Cambridge, 1985.
4. N.P. Ong, "A simple discussion of the Berry phase," Key Article: http://physics.princeton.edu/~npo/SurveyTopics/Berryphase.html
5. R. Bishop and S. Goldberg, Tensor Analysis on Manifolds, Dover, 1968, 1980 (e.g., p. 222).
6. **MTW**: Misner, Thorne, Wheeler, **GRAVITATION**, Freeman, 1970-1973 [includes discussion of curvature 2-forms, e.g., p 350]. This is sometimes referred to as "The Telephone Book" because of its large pages and thickness.

7. Steven Weinberg, Gravitation and Cosmology, Wiley, 1972 (e.g., p111).

8. Chapter 11, "Differential Calculus on Manifolds." www.cns.gatech.edu/~predrag/courses/PHYS-6124-12/StGoChap11.pdf
9. Introduction to Tensor Calculus for General Relativity, MIT 1999 Edmund Bertschinger.
10. William Baylis, Electrodynamics: A Modern Geometric Approach, Birkhauser, 1999.
11. Theodore Frankel, The Geometry of Physics, 2nd Ed.,Cambridge 1997, 2004.

12. Orlin Stoytchev, "From Foucault's Pendulum to the Gauss-Bonnet Theorem," arXiv:1701.01666 6 Jan 2017.

13. S. Carroll, "Lecture Notes on General Relativity," https://ned.ipac.caltech.edu/level5/March01/Carroll3/Carroll3.html
14. Robert M. Wald, General Relativity, Chicago, 1984.
15. Abraham Pais, Subtle is the Lord, Oxford, 1982
16. https://physics.stackexchange.com/questions/77973/riemann-tensor-in-2d-and-3d
17. http://www.physicspages.com/2014/10/21/ricci-tensor-and-curvature-scalar-for-a-sphere/
18. "Cartan's Structural Equations and the Curvature Form," Symposium, mathstat.carleton.cal/~ckfong/s43.pdf

19. www2.warwick.ac.uk/fac/sci/physics/current/teach/module_home/.../lecture12.pdf
20. https://einsteinrelativelyeasy.com/index.php/general-relativity/61-the-riemann-curvature-tensor many topics available under Einstein relativity easy including "Introduction to Covariant Differentiation."
21. Robert T. Jantzen. "dr bob's elementary differential geometry," 2008. www.win.tue.nl/~rvhassel/Onderwijs/.../Examples-diff...diff.../dr_bob_dgn2010.pdf
22. https://math.stackexchange.com/questions/410844/curvature-tensor-of-2-sphere-using-exterior-differential-forms-tetrads
23. Lctr-cfrm1.pdf PHYS 4390: "General Relativity non-coordinatge basis approach."
24. Michael Nielsen, "An introduction to Yang-Mills theory," Chapter 5 at - michaelnielsen.org/blog/yang_mills.pdf.

Additions: **"Intrinsic geometry"** uses math that can be expressed using a metric line-element or "first fundamental form:" smooth curve lengths (dating back to Euler 1736), areas, torsions, connections, Gaussian curvature or curvature tensors (e.g., Riemann 1854). "Extrinsic geometry" needs an immersion or embedding of a manifold in a space of higher dimension. The "shape operator" for example uses normals to a surface (needing a dimension perpendicular to the surface). There was supposed to be a difference between
extrinsic curvature of objects embedded in a space and intrinsic curvature using lengths of curves in a Riemannian manifold. But **Gauss' Theorema Egregium** (Outrageous Theorem) says that the same K pertains to both.

A "frame field" (is a set of n vector fields) as an orthonormal basis for a tangent space -- and has direct physical meaning. What are called "coordinate bases" are like partial derivatives and are not unit vectors for frame fields.

Consider a smooth curve. Zoom in on a point and look at its neighborhood. The exterior derivative of a 0 form measures the difference between the zero forms evaluated at points infinitesimally apart. As the points draw nearer and nearer, the difference becomes more and more a differential. This physically measures the difference in flow that exits one point and enters the other. If one were to add all these infinitesimal differences across a curve, the interior regions cancel leaving the exterior difference behind. For when the flux exits one end, it immediately enters the next.

Exterior derivative generalizes gradient of a scalar function, and Stoke's theorem generalizes the fundamental theorem of calculus (FTC).
If f(x) = dF/dx, dF = fdx, $\int$ dF = F = $\int$ fdx $\rightarrow$F(b)-F(a) at **ends** of a line a to b, i.e., ∂ = . . At x=a and b. FTC. fdx is a 1-form and F is a 0-form (just a function).
Stokes Theorem (1854) now: If dω is an n-form, $\int_\Omega d\omega = \int_{\partial\Omega} \omega$
Similar to $\int_S (\nabla x A)\cdot n dS = \int_{\partial S} A\cdot dr$. A·dr is a 1-form.

The name exterior derivative because the integral of the "exterior derivative" leaves the difference of the exterior behind. Loosely the word "integral" cancels "derivative" leaving " exterior" behind...hence its name.

Word Option: å ,ß=opt s, ∂=opt d ∆=opt j √=opt v ∫=opt b. µ=m, ö u, Ω=z, and ˚ for temperature degrees (not Kelvins).
LaTeX var___ =ϱ , ϑ , φ and varsigma=ς .

The Lie Derivative

Dave Peterson, 3/18/19 – 6/22/19

A vector "Lie derivative," $\mathcal{L}_X \mathbf{Y}$, evaluates the change in a vector field, Y, subjected to a flow defined by a separate vector field, X. The overall coordinate invariant concept of Lie derivatives is more general than this and applies to any differential manifold for scalar functions {f}, vector fields {**X**}, one-forms {ω}, or tensor fields {T}. The concept dates back to 1931 but is based on the earlier works of Sophus Lie near 1890. "**X**" as a contravariant "vector" is also a differential operator on something: for example, vector $X = X^i(\partial / \partial x^i)$ = contravariant component times a "coordinate basis." One expression for the Lie derivative at a point p with respect to two vector fields is called the commutator, $\mathcal{L}_X \mathbf{Y}_{(p)} = [X,Y]_{(p)} = (XY\text{-}YX)_{(p)}$, and using it for calculations is straightforward. But an equivalent form is in terms of a more traditional limit of difference ratios that requires first finding integral curves for vector flows to move points and tangents on a manifold. It is unusual in using two different mappings together: a displacement map for points (say "q" $\mapsto \varphi(q)$) and a tangent vector map, φ_*, for returning a vector back to location q [See Fig 1].

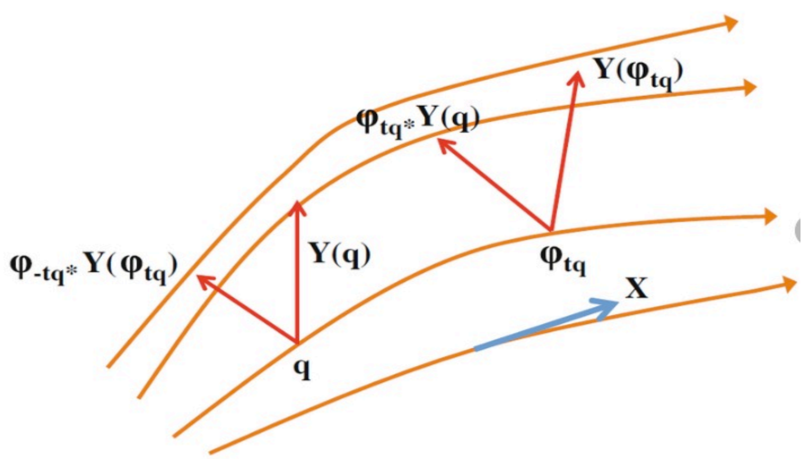


Figure 1. Flow φ moves point q along a streamline of X and can relocate vector Y. φ_* is a tangent mapping that moves tangent vectors and can move $Y(\varphi_{tq})$ back to its starting point for comparison. [Lim Zheng Liang, Singapore].

OUTLINE:
1. Lie Derivative of a scalar function, Directional derivatives.
2. Lie Commutator [X, Y].
3. Lie Derivative of a vector field, PushForwards, PullBacks, forms.
4. Flow streamlines and integral curves
5. **Background Material and definitions in Differential Geometry.**
6. Applications of Lie Derivatives
7. References.

Simplest Case: **The Lie derivative of a scalar function, f:** which is just the familiar "directional derivative" at a point p -- the scalar value of a gradient vector projected on a vector of a vector field: [survey of definitions]

[Eqn. 1] $\mathcal{L}_X \mathbf{f(p)} = (\mathcal{L}_X f)(p) = X(f) = X[f] = X\cdot\nabla f(p) = (X\cdot\nabla)\, f(p) = \partial_X f =$

$X^{\mu} \partial_{\mu} f(p)= (d/dt)(f(p+tX)|_{t=0} = \partial_{t} (f(P(t,p))$

[-- with a variety of different notations and conventions across sources]. If this Lie derivative is zero, then its scalar function has the same value all along its streamlines or "integral curves" (see function examples below). { f can be treated as a scalar function or as a 0-form}.

Note that elementary introductions to directional derivatives often begin with familiar "**unit** vector" "physical" bases (like u_{μ} or e_{μ} = i, j, k for $\Re^{3}$) combined with vector values coefficients $\mathbf{a}_{\mu}$. So, in calculus, the directional derivative was presented as $\mathbf{v}\cdot\nabla f = (\mathbf{v}\cdot\nabla)f$. But in differential geometry, a preferred basis may instead be a tangent vector basis like $\partial_{x} \equiv \partial / \partial x$ (which are called "coordinate bases") with vector $X = X^{i} (\partial / \partial x^{i})$. An example is angle phi-base $\partial_{\phi} = r \sin\theta \; e_{\phi}$ for spherical polar coordinates with unit vector bases. Or, in terms of the metric tensor, $\partial_{\phi} = (\sqrt{g_{\phi\phi}})\, e_{\phi}$. These ∂_{x}'s are more than basis names—they function as first order differential operators.

A vector can be written in three different ways: $\mathbf{A} = \mathbf{a_{\mu} e_{\mu}} = \mathbf{A^{\mu}} \partial_{\mu} = \mathbf{A_{\mu} e^{\mu}}$ [Eqn. 2]. (with "Einstein convention" of summing over all repeating index values μ = 1,2,3… together). A^{μ} is called contravariant (with tangent vector bases) and A_{μ} is called covariant on gradient coordinate curves: $e^{\mu} = \nabla u_{\mu} = \mathbf{d}x^{\mu}$ of size $|1 /(\sqrt{g_{\mu\mu}})|$ – usually NOT unit vectors. The product of bases $(\partial_{i})(\mathbf{d}x^{j}) = \delta_{ij}$ (=1, or 0 if $i \neq j$).

Example [Burke]: the derivative of $f = x^2+y^2$ in the V direction
$(\partial / \partial x + \partial / \partial y)$ {= (1i+1j)·∇} is given by $V[f] = (\partial / \partial x + \partial / \partial y) (x^2+y^2) =2x+2y$.

So the funny notation $\mathbf{v}$(f) or $\mathbf{v}$[f] makes sense because it is an application from the left that pre-empts having to formally take a gradient ∇f.

As another example [Wik], if vector field $X = \sin x \, \partial_{y} - y^2 \partial_{x}$ and scalar field $f = x^2 - \sin y$, then $\mathcal{L}_X f = X\cdot \mathbf{d}f = (\sin x \, \partial_{y} - y^2 \partial_{x})(2x\mathbf{d}x - \cos y \mathbf{d}y) = -\sin x \cos y - 2xy^2$. [using $\partial_{x} \mathbf{d}x = 1$ -- contravariant base of a covalent base is δ_{ij}. "**d**" is "exterior derivative," and dx is a "1-form" and **d**f is somewhat like a gradient, ∇f].

For the scalar function case, the terms Lie derivative, covariant derivative, and directional derivative all coincide. For a scalar function all by itself, covariant differentiation is just partial differentiation: $f_{;a} = \partial_{a} f$.

The Commutator, [X,Y] (this will be used for $\mathcal{L}_X \mathbf{Y}_{(p)}$ below):

Since a vector field X is a differential operator that can operate on a function to yield another function, a second vector field Y can operate on that function. So, $Y\{X[f]\} = \partial_{Y}(\partial_{X} f)$ [MTW]. In general, a commutator of such double applications will not commute. In coordinate-based calculation (notation $_{,\alpha}$ = derivative wrt x^{α}):
$[X,Y] f = X^{\alpha} \partial / \partial x^{\alpha} (Y^{\beta} \; \partial f/ \partial x^{\beta}) - Y^{\alpha} \partial / \partial x^{\alpha} (X^{\beta} \; \partial f/ \partial x^{\beta}) = [(X^{\alpha}Y^{\beta}{}_{,\alpha} - Y^{\alpha}X^{\beta}{}_{,\alpha})(\partial / \partial x^{\beta})] f$.

The presence of the $(\partial / \partial x^{\beta})$ basis says that this commutator [X,Y] is itself a contravariant vector field. {see definitions in Background section}.

If vectors X and Y are themselves just simple contravariant bases (or covariant dual bases), then they do commute: $[\partial / \partial x^{\alpha}, \partial / \partial x^{\beta}] = 0$ and $[\mathbf{e}_{\alpha}, \mathbf{e}_{\beta}] = 0$. [MTW].

Later on, we will consider a single basis $\partial_t = \partial / \partial t$ for transport along a tangent with parameter t such as time (streamline flow).

Many texts describe how to picture such commutators as approximate parallelograms that don't quite close up but have a gap $\simeq$ [X,Y] when vectors do not commute. That is, start at a point p_o and draw vector arrows $X(p_o)$ up to a new p_1 and $Y(p_0)$ across to a new point p_2. Then draw more side arrows again from these points: $X(p_1)$ and $Y(p_2)$. Do the final vector tips match up or not?

Roger Penrose says that the Lie bracket measures the gap in an incomplete quadrilateral of arrows made alternately from ϵX and ϵY vectors (epsilon is a tiny number resulting in a tiny length). He adds that $\mathcal{L}_X Y$ states how the vector field Y actually changes as contrasted with just having it "dragged along" by X [Penrose]. If [X,Y] = 0, a function or "flow" phi ϕ_t pushes a point along X; and flow theta θ_t could push a point in the Y direction. Then for a point p, $\theta_{-t}\, \phi_{-s}\, \theta_t\, \phi_s(p) = p$ – a curvy parallelogram with sides determined by small s and t and having no gap. But, if $[X,Y] \neq 0$, then the parallelogram won't close and will have a gap ~ st [X,Y](p) [Frankel].
{Elementary examples of torsion include the helix curve and the "Archimedes" screw surface}.

If the commutator was written for 3-vectors in $\Re^3$, then the vector [X,Y] = $(X\cdot \nabla)Y - (Y\cdot\nabla)X$. When this equation holds, the Lie derivative is said to be "torsion free."

The Lie derivative of a vector field Y with respect to a vector field X at a point p can be defined as:

$$\mathcal{L}_{\mathbf{X}}\mathbf{Y}(p) = [X,Y](p) = \partial_X Y(p) - \partial_Y X(p), \text{ [Wik] \qquad [Eqn. 3].}$$

where the commutator [X,Y] is called a "Lie bracket" and is itself a vector field. For notation, the capital X in ∂_X means $(X^j \partial_{xj})$ for all its coordinates. Initially, this commutator-bracket definition looks new and strange, so lets postpone it momentarily until we can derive it in a way that agrees with our previous understandings and intuition about derivatives.

In ordinary calculus, we define a derivative as a limit of difference ratios

$f'(x) = df/dx = \mathrm{Lim}_{\Delta x \to 0} \{\Delta f/\Delta x = [f(x+\Delta x)-f(x)]/ \Delta x \}$.

We want something similar to this.

For the Lie derivative we can state as difference ratios [Frankel] :

$$\mathcal{L}_{\mathbf{X}}\mathbf{Y} = \lim_{t\to o} [Y_{\phi x} - \phi_{t*}\, Y_x]/t = \lim_{t\to o} [\phi_{-t*} Y_{\phi x} - Y_x]/t. \qquad \text{[Eqn. 4]}$$

We move the vectors Y in a special way and have to compare differences at a common point like x = p(t=0). In **Figure 1** above, we can pick the common point either as p or as $\phi(x,t)$ slightly ahead of p. Here the first difference in equation 4 is at the point $x + \epsilon$ pushed slightly to the right, and the second expression is at $x_o = x(t=0)$. In the limit,

both expressions are equivalent. We haven't said yet what ϕ or ϕ_{t*} mean. So, to go further, we first need to present some background and introduce new symbols to develop the Lie derivative from this definition.
{Some of this background is also given in the later section "Some Definitions."}

Equation 4 is itself an expansion of another **definition** $\mathcal{L}_X Y_p = (d/dt)|_{t=0} (\varphi_{-t*})Y_{\varphi t(p)}$ at point $\varphi(t,p)$. $\varphi(-t, \varphi(t,p)) = \varphi(-t+t, p) = \varphi(0,p) = p$. φ_{-t*} is a "push-forward" that happens to be pushing backwards –t.

Mappings and Push-Forward Tangent Mappings:

Tangent vectors are defined by components $v^\alpha = dx^\alpha/dt$ (α = 1,2,…m, the dimension of M if we have a mapping from manifold M to N; coordinates x in M and y in N). A corresponding pushed ahead tangent vector in N is

$dy^\beta/dt = (\partial y^\beta/\partial x^\alpha)(dx^\alpha/dt)$ where α and β are indices over the dimensions of the manifolds. The middle term linear transformation is called the "**Jacobian**" or Jacobian matrix $\mathbf{J} = J_{\alpha\beta}$. If N = M so that both manifolds have the same dimension, then the middle term is a square matrix (with a square determinant also called the Jacobian determinant). **$J_{\alpha\beta}$ is a "tangent map" F_* from v to $F_*(v)$** {using mapping label "F" here because we are referring to coordinates rather than parameterized flows. {The asterisk $*$ (unless placed high *) denotes transport of tangent vectors rather than just movement of points}.

Note that a push-forward of a vector is different and obeys:

$\partial_x \rightarrow (\partial F^\alpha/\partial x)\ \partial/\partial \eta^\alpha$. But for F: $R^m \rightarrow R^n$, eta in N is just x or y (or z).

Math books emphasize that a tangent vector is "an equivalence class of curves" [Burke] and so also are basis vectors like $\partial_x = \partial/\partial x$.

Example 1 F: $R^2 \rightarrow R^2$ $\mathbf{(x,y) \rightarrow F(x,y) = (f_1,f_2) = (x^2 - y^2,\ 2xy)}$ [O'Neill p.36]
(called the "realification of the complex square function," and corresponds to a vector field flow $V = 2\,\partial_\theta$)

{e.g., p = (a,b) = (3,1)→ p'=(8,6), and (0,0)→ (0,0)}.

The mapping F is a non-linear transformation: matrix M: $f_i = M_{ij}\, x_j$.
Let vector v = (v_1,v_2), and J = Jacobian transformation. Unit base vectors i and j.
$F_*(v) = F(p+tv)'(0) = v\cdot(\nabla f_1, \nabla f_2) = (v_1 i+v_2 j)\cdot(2xi-2yj, 2yi+2xj) = 2\,(xv_1-yv_2,\ v_1 y+xv_2)$.
Initial velocity $df_i/dt = J\,(dx^j/dt) = (\partial f_i/\partial x^j)(dx^j/dt)$ [as 2 by 1 column vectors]

i.e., $\beta^1 = (df_1/dt) = 2x(dx/dt) - 2y(dy/dt)$, and $(df_2/dt) = 2y(dx/dt) + 2x(dy/dt)$.

J can be called "dF."

So, for small increments ,

$F(a+\epsilon_1, b+\epsilon_2)|p \simeq F(p)+J d\mathbf{x} = F(a,b) + (2a\epsilon_1 - 2b\epsilon_2 ,\ 2b\epsilon_1 + 2a\epsilon_2)$.

{check: $F(1+0.01, 1+0.01) \simeq F(1,1) + 0.01(2-2, 2+2) \simeq 2.04$ vs. Direct F = 2.0402 }.

{Note: Taylor approximations aren't relevant because they apply to scalar functions. Here we do need the Jacobian, J }.

A transformation "T" is **linear** if: T(cv) = cT(v) and T(v+w) = T(v) + T(w).
The function for incremental change $dF = J d\mathbf{x}$ (or $dF = J d\boldsymbol{\epsilon}$) is linear because if we let $dx \rightarrow dx+\delta x$ and $dy \rightarrow dy+\delta y$, then we do get $dF' = J d\mathbf{x} + J\delta\mathbf{x}$,

and for c$d\mathbf{x}$ we have dF'= cJd$\mathbf{x}$.
So, in the close vicinity of a given point, p = (a,b), we have a **linear transformation** away from that point and can apply linear algebra.
The Jacobian matrix in this example is [2x -2y; 2y + 2x] and has det J = $4(x^2+y^2)$ which is positive except for the origin, (0,0) , i.e., on $\Re^2 \setminus \{(0,0)\}$. The inverse matrix then exists and is
$J^{-1}= (1/[2(x^2+y^2)])$ [x +y; -y x], and d$\mathbf{x}$=J^{-1} d$\mathbf{f}$.

So, is F a diffeomorphism? Well, not quite because it is not one-to-one (bijection). That is, F(1,0) = F(-1,0) = (1,0); two points map to one point. Actually it is a whole family of parabolas: F(± x, 0) = (x^2 ,0), and a bit of y shifts the parabola up or down. But, we can still discuss "local diffeomorophisms" at infinitesimally small distances from a reference point p.

In retrospect, the flow streamlines in this case come from a simple vector field similar to the common circular flow V= $\partial_\theta = \partial / \partial \theta = y\, \partial_x - x\, \partial_y$. But this flow is faster and could be called V = 2 ∂_θ. The time parameter integral streamline curves turn out to be $F_t = (r^2\cos^2 t - r^2\sin^2 t,\ 2r^2 \sin t \cos t) = (r^2\cos 2t,\ r^2 \sin 2t)$ that goes twice around a circle. [END].

Example 2 F: $R^2 \rightarrow R^3$, Plane to Spherical Cap: Project a circular area on a plane up to an upper half-sphere. That is, find a map (F) and its Tangent Mapping (F_*) from a u-v plane in R^2 to a unit radius spherical cap with coordinates x,y,z in R^3:

Let u and v be orthogonal axes for R^2 used within a unit radius circular domain on the plane from a center point p_o = (0,0). Suppose we consider coordinate map is F(u,v)→ (x,y,z) to the spherical surface directly above the plane. So (u,v) → (u, v, z) where $z =(1- u^2- v^2)^{½}$.

Examine a sample point in the plane at p = (0.5, 0.5) so that z = 0.707 . That is, q = F(p) = (0.5, 0.5, 0.707) in R^3.

Let the initial partial velocities in the u and v direction of the plane be unity so that we look at a tangent vector V_p = (1, 1) and push this forward to the cap-surface via a tangent map $F_*(V)$. What are the partial speeds x_u (holding v constant) and x_v (holding u constant) on the cap?

We can have the tangent map on S^2 be derived from a straight line curve α(t) from (0,0) to a great circle curve β(t) on S^2. Use $\partial z/ \partial u = -u/z$ – speeding forward in u and v makes z-speed move downwards.

β '(t) =F_*(α'(t)) = F_*(V) =V[F] where $\beta_j(t) = \Sigma(\partial F_i / \partial x^j)(p)$ times base units at F(p) – namely i,j,k natural or unit vectors. OR, β'(t) = J α'(t) where J is the Jacobian Matrix = [1, 0, -u/z ; 0, 1, -v/z] times a column matrix [i; j; k] . So an initial u unit vector in R^2 will yield (i – 0.707 k) at point q. And a v unit vector will yield (j – 0.707k). So V_p =(1i+1j) → (i+j -1.4 k) at point q. The absolute value speed goes from the original $\sqrt{}2$ to speed 4 on S^2. It is ok to have a faster speed since we mapped from a flat surface to a spherical surface having longer arc lengths.

Alternatively, we could have used a "geographical patch" for Earth coordinates with equator at north-south angle θ = 0 and longitude lines for ϕ [-π , + π]. Then
$\mathbf{x}$(u,v) = F(ϕ, θ) = (x,y,z) = (r cos θ cosϕ , r cosθ sinϕ , r sinθ) [e.g., O'Neil p140].
Then partial speeds are F_u= $\mathbf{x}_u$(u,v) = r(-cosθsinϕ , cosθ cosϕ , 0), so $|F_u|$ = r cosθ .
And F_v= F_ϕ = r(-sinθcosϕ, -sinθ sinϕ, cosθ), $|F_v| = r\Theta^{hat.}$

Here, theta-hat is a unit vector upwards. The view here is for uniform angular speeds for θ and ϕ angles on a sphere.

The previous example was uniform radial speeds for u and v.

Example 3 . [Benn, p129] $M=\Re^2 \rightarrow N = \Re^3$, with mapping

$p=(x,y)\mapsto (f^1,f^2,f^3)= (x^2, xy+1,y)$.

There are several approaches to pushing a vector $V=(v^1,v^2)$ to new tangent vectors in N (note upper case for contravariant tangent vectors). One is just working out the Jacobian matrix $[\partial f^i/\partial x^j]$ for two rows of j :($\partial/\partial x$, $\partial/\partial y$) and three columns of i for f^1, f^2, f^3. The "J" matrix for row 1 is (2x, y, 0) and row 2 is (0, x, 1). Multiply by row (v^1,v^2) from the left yielding the answer: $D_V f(p) = (2v^1x, xv^2+yv^1, v^2)$ – a new 3-vector.

Another is to find the directional derivative using a 3d $V\cdot\nabla f = (v^1,v^2,0)\cdot$ 3d ∇ gradient on function f. This is equivalent to the Jacobian J above.

The original **definition** of the directional derivative of a function, f, is:

$$D_V f(p) = \mathrm{Lim}_{h\to 0}\,[(f(p+hV)-f(p)\,)/h]$$

The f(p+hV) portion = $((x+hv^1)^2, (y+hv^2)(x+hv^1)+1, (y+hv^2))$.

Expanding and subtracting f(p) and chopping off h's (and ignoring any tiny h^2's) gives the same result as above.

In the following, it might help to keep symbol names somewhat consistent. As a convention, we might prefer to use the symbol "**F**" above for mapping using coordinates versus $\phi = \phi(t) = \phi_t$ for mappings using time flows along streamlines (ϕ_t moves a point x forward by t seconds). {But, different sources use different conventions}.

In general, let: $\{ t\in I \subset \Re \},\ \ \gamma : I \rightarrow\ \mathbf{M} \text{ --}\phi \rightarrow \mathbf{N} \text{ -- } f \rightarrow \Re.$

Parameter $t\in$ real-interval I (not necessarily time). M and N are differential manifolds possibly of different dimensions, but often here N=M. The symbols ϕ or F refer to a smooth mapping of **points** x from M to $y\in$ N, and scalar function f is any function from N to real numbers (or, M to $\Re$). Associated with a linear transformation mapping ϕ are two maps: the "Push Forward" $\phi*$ or ϕ_* for **tangent vectors** in M at point p to tangent vectors in N at p' . The push-forward for F essentially looks at changes to $F = (f_1, f_2\ldots)$ using the Jacobian, dF =Jdx for tangent-mappings. We might think of an integral curve $\phi*$ as similar to $d\phi = (d\phi/dt)dt$ on tangent vectors along streamlines.

Consider a curve γ (t) in M. ϕ_* is a linear transformation mapping of a tangent vector in M to another corresponding tangent vector in N. For short parameter increments s and t, $\phi_*(s+t) = (\phi_* s) + (\phi_* t)$. A curve γ can be pushed forward by composition "∘ ": $\gamma_* = \phi \circ \gamma$. Summarizing, X and Y are vector fields with ϕ(p) and ψ(p) as respective integral curves starting at point p, and ϕ_t and ψ_t the associated diffeomorphism [Benn].

To make this more conventionally clear, consider the parameterized curve gamma = $\gamma = x^\alpha(t) = (x_1(t), x_2(t), \ldots)$ and "push it ahead" with a mapping ϕ from M → N to a new relocated curve labeled by $\gamma_* = y^\beta(t) = y^\beta\,(x^\alpha(t))$ – just a composition. So, $\gamma_* = \phi\circ\gamma$

Pullbacks *:

[Frankel] For a function f on $y \in N \to \Re$, $(F^*f)(x) = (f \circ F)(x) = f(y(x))$ where F^* is called a "pullback" and simply means that the function expressed as f(y) is referenced back to $x \in M$: f(y(x)).

{*"Asterisk-high"* F* means reference backwards, F_* or F_* means push forward from manifold M to N}.

A vector V acting on the pull-back of a function f is $V(F^*f) = V[f(y(x))] = (F_*V)(f) = df(F_*V)$.
{For 1-forms ω acting on vectors, $(F^*\omega)(v) = \omega(F_*v)$, $(p,V) \to F_{*p}$, $D_VF(p) = dF(p)V$.}
For the case of a covariant tensor (a p-form, α_p, the general pull-back is expressed as:
$F^*\alpha_p(v_1, v_2, v_3, \dots v_p) = \alpha_p(F_*v_1, \dots, F_*v_p)$ [Frankel] – a distributed push-forward of p tangent vectors.

These concepts will carry over to flows, ϕ_t.

Comment Push forward (tangent map) &, Pull Back: F* pulls back forms ω. Forms ω are functions on vectors V or v, F_* pushes v's. (Contravariant) Vectors are operators on functions f.

Useful Equations: For a Mapping F from manifold M to N, let ψ or ω be one-forms in N. X or V are vector fields. γ is a parameterized curve. ∘ is composition of functions. x^i are variables in manifold M, and η^i or y^i are coordinates in manifold N (y's are overused and can be confusing). p is a point in M. {u is a patch variable like x}.

$F^*\psi(V) = \psi(F_* V) = (\psi \circ F)(u) = \psi_{F(u)}(dF_u(V))$ $\{F_* \sim dF_u\}$
$(F^*V) f = [(F\circ \gamma)_*(o)] f = V(f \circ F) = \gamma_{*o} f \circ F$. [Bishop p55] for curve γ, $\gamma_*0 = V$.
$V(F^*f) = V[f(y(x))] = (F_* V)(f) = df(F_*V)$
For Components: $F_* X_p = X^i(\partial / \partial x^i)F^j(\partial / \partial \eta^j)|_{q = F(p)}$. [Benn, p 148]
Or for covariant base: $dx^j(F_{*p} X_p) = X^i_p(\partial F^j/\partial x^i)_{(p)} = (\partial F^j/\partial x^i)_{(p)} dx^i(X_p)$
For $\omega = \omega_j dx^j$, $F^*\omega = (\omega^j \circ F)(\partial F^j/\partial x^i) dx^i$. [Benn p 149]
Jacobian J: $dy^\beta/dt = (\partial y^\beta/\partial x^\alpha)(dx^\alpha/dt)$. { v' = J v}. $^*\omega_\alpha(p) = (\partial y^\beta/\partial x^\alpha)\omega_\beta(p')$
{Integral Curves} $\mathbf{d\phi(t)/dt = V(\phi(t))}$, **or** $\mathbf{dx^i(t)/dt = V^i(x^i)}$ **}.**

Flows: (Streamlines. Address the Lie derivative as difference ratios). The Lie derivative could be called a "flow" derivative.

Consider a given vector field X to be generated by a "**flow**" such as the velocity field, $v^j(x) = dx^j/dt$, analogous to a streamline flow of water as a function of time, $\mathbf{v}_p = d\phi_t(p)/dt|_{t=0}$ [Frankel] . X could be the wind vector field above an ocean. Another field Y can also be generated from its own scalar function, ψ(t) . At parameter t = 0, we have tangent vector X along its streamline and vector Y(x) along its streamlines. $x = \phi_t t$ is a point t seconds along the streamline curve of X. Let the vector field Y have direction at angles away from X and also be carried along with the flow from some initial point $p = x_o$ to a later point $x(t) = \phi(t)x_o$. For "curvy" flows, the angle of Y with respect to X may be quite different from point to point. An integral curve is a parameterized trajectory or streamline through a given point.

In words, what the "difference ratios" expression for the Lie derivative is saying (equation **4** above) is: For a vector Y at x_o = p, evaluate a new Y at p'=x(t) at a point t seconds along the X streamline using ϕ_t. Then use "tangent-mapping" to map new vector Y backwards using ϕ_{-t*} to the original point p and compare it to the original Y.

{The essence of this can also be stated in a variety of different ways" At new reference point p' = ϕ_t(p), find Y there and subtract from it the original Y_P tangent-mapped to p' using push forward: $\phi_{t*} \sim d\phi_t$. Notice that eqn 4 has these two equivalent expressions. And then two other forms can be stated using pull-backs (and a variety of texts reference one or two of these four forms) }.

Finding Integral curves:

Find a mapping ϕ: $t \in \Re \rightarrow \Re^2$ such that **$d\phi(t)/dt = V(\phi(t))$,** and p=ϕ(0) = x_o for a vector field with components V = (v_1, v_2) on $\Re^2$. ϕ = (f_1, f_2) or (ϕ_1, ϕ_2) or (x(t), y(t)) functions of parameter t (different sources have different names—but it is easiest to think of $\phi = \phi(x^1(t), x^2(t), \ldots x^m(t))$. Then the first order differential equations look like:

$dx^i(t)/dt = V^i(x^i)$

treating the left side as a function of t but right side as a function of x.

Case 1: For a constant Field Let v_1=1, v_2 = 2, ϕ(t) = (0,0) + $\int$(1,2)dt = (1t, 2t). And $d\phi(t)/dt = [d\phi^1/dt, d\phi^2/dt]$ = (1,2). The "streamlines" here are straight lines of slope 2 one of which passes through the origin. {e.g., " http://planning.cs.uiuc.edu/node382.html }.

Case 2: Linear Velocity Field, **X** = $-2\,\partial_x - 1\,\partial_y$ = (-2, -1): So, $df_1/dt = -2f_1$, $df_2/dt = -f_2$. So, ϕ(t) = (0,0) + (exp(-2t), exp(-1t)).

Here, ϕ(0) = (1,1). $d\phi/dt|_o$ = (-2, -1) so the initial slope is ½. For tiny times, $\phi(\epsilon) \simeq (1-2\epsilon, 1-1\epsilon)$.

Case 3:

Suppose vector field **X** = $(x\,\partial_y - y\,\partial_x) = (-y, x)\cdot(\partial_x, \partial_y)$ starting at p = (a,b). Note that for Polar coordinates (r, θ), position **r** =i x+j y = i rcosθ +j rsinθ and Theta θ =$\tan^{-1}$(y/x). Suppose |r| = constant = unity. $\nabla\theta$ = -i y +j x = (-y,x) =e_θ.
So, field **X** = $e_\theta = \partial_\theta$ positive CCW rotation on the unit.
Solve **integral curves** such that: $\partial\phi_t(p)/\partial t = X(\phi_t(p))$ {X and ϕ_t describe a rotation}.

Let ϕ = (ϕ_1, ϕ_2) obey $\partial\phi_1/\partial t = -\phi_2$ and $\partial\phi_2/\partial t = +\phi_1$ [Benn p158].
{or more clearly, dx(t)/dt = -y and dy(t)/dt = x}. For operator D = $\partial/\partial t$, we have the coupled equations $D\phi_1 = -\phi_2$ and $D\phi_2 = +\phi_1$.
Applying a second derivative operator : $D^2\phi_1 = -D\phi_2 = -\phi_1$ and $D^2\phi_2 = D\phi_1 = -\phi_2$ for the form $(D^2+1)\phi_i = 0$. Both sin t and cos t are solutions of this equation so that both ϕ_1 and ϕ_2 have solutions of the form $\phi = c_1\sin t + c_2\cos t$.

Recall that the rotation matrix has (a';b') = [cosθ , sinθ ; -sinθ ,cos θ](a;b).
p' = Mp. Rotating a position vector p CCW is the same as rotating axes **backwards** ClockWise, CW.
Write: (a'; b') = (a cosθ +b sinθ , -a sinθ +b cosθ).
Let θ grow with time, θ = t.
So, the diffeomorphism ϕ_t has [e.g., Benn, p 153]:

p=(a,b)→ (a', b') = ϕ_t(a,b) = (a cos t + b sin t, b cos t - a sin t).
With respect to the above coupled equations, this solution implies t→ -t !
This will be used in a later problem [i.e., Find the Lie derivative for the above X field and a given Y, Example 2 below], dϕ/dt$|_o$ = (-b,a), so the initial slope is –a/b.

Case 4 Example: "Consider the quadratic vector field on $\Re$," [Frankel exercise p 35] **V(x) = x^2 d/dx**. From the requirement dϕ(t)/dt = **V**(ϕ(t)) = **V**ϕ(x) treated as a derivative of **x** for ϕ = x(t). Solve dx/dt = x^2 with x(0) = p =x_o. The time derivative on the left is replaced with an operation by V on the right!

Integrate $\int dx/x^2 = \int dt$ with limits 0 to t and p to p(t). Result is $\phi_t = x(t) = [x_o/(1-x_ot)]$, and ϕ_{t*} =dx(t)/dt = V(x(t)) = $x(t)^2 = [x_o/(1-x_ot)]^2$. And this "push forward" satisfies dx(t)/dt = V(x(t)). That is, $(d/dt)[x_o/[1-x_ot]\,) = x_o^2/(1-x_ot)^2 = \mathbf{V}\phi(x)=(x^2\,\partial_x)\,\phi(x)=x^2\,\partial_x(x/[1-xt])|_p$.
{For small t and p, ϕ_t(p=0)= tV, and here $\phi_t \sim x_o+x_o^2t$, but V=x^2, so yes ~ tV }.
{Note, for example, Claim by [Burke p. 124] }.

Case 5: Flow Field [Burke p 94]
Let V = y ∂_x –(y + x) ∂_y and $\phi = (\phi^1, \phi^2)$. Then $\partial\phi^1/\partial t = \phi^2$, $\partial\phi^2/\partial t = -[\phi^2 + \phi^1]$.
{or more clearly, dx(t)/dt = y and dyi(t)/dt = -[y+x] }. Take another time derivative $\partial/\partial t$ of the first equation and plug it into the second equation.
Then $\partial^2\phi^1/\partial t^2 + \partial\phi^1/\partial t + \phi^1 = 0$,or $(D^2+D^1 + 1)\phi^1 = 0$. From mechanics, "The integral curves are spirals representing damped harmonic motion."

$\mathcal{L}_X$ Y = [X, Y]:

Now, to derive the **commutator** form [,] for the Lie derivative:
$\mathcal{L}_X$ **Y** is a vector operator, so let it operate on any scalar function f(x) over a "small" (linear) neighborhood of x. Let X(f) be a differentiable function g_o(x) and f(ϕ_t,x) = f(x)+tg_t(x) so that composition (f ∘ ϕ_t) = f + t g_t(x) where g_t = g(t,x). **[Eqn. 5].**

This is like a Taylor's series expansion: f ∘ $\phi_t \simeq$ f + tXf + $t^2/2$ X^2f + … and just keep the first two terms for linearity {the existence of function g(t,x) is called "Hadamard's Lemma" [Frankel, p126].}

Now apply Eqn. 4 to scalar function, f, from M to $\Re$ (and let manifold N = M) .
From equation 4, $[\mathcal{L}_X \mathbf{Y}](\mathbf{f}) = \lim_{t\to o}(1/t)[Y_{\phi x} - \phi_{t*}\, Y_x](f)$.
But, $[\phi_{t*}\, Y_x](f) = Y_x(f \circ \phi) = df(\phi_*\, Y_x) = Y_x(\phi^*f)$ {and we don't have to explicitly use the "pull-back ϕ^*}. {some texts emphasize the use of pullback}. So, using Eqn. 5, we now have:

$[\mathcal{L}_X \mathbf{Y}](\mathbf{f}) = \lim_{t\to o}(1/t)[Y_{\phi x}(f) - Y_x(f\circ\phi_t)] = \lim_{t\to o}(1/t)[Y_{\phi x}(f) - Y_x(f + tg_t)] =$
$\lim_{t\to o}(1/t)[Y_{\phi x}(f) - Y_x(f)] - \lim_{t\to 0} Y_x(g_t)] =$
$X_x[Y(f) – Y_x(g_o) = X_x[Y(f)- Y_x[X(f)]$, So, $\mathcal{L}_X$ **Y = [X,Y],** Eqn 3 again.

A variety of these derivations can be found elsewhere [e.g., Burke].

In local coordinates, the commutator can be re-written as:

$$[X,Y]^i = XY-YX = \sum_j \left\{X^j(\partial Y^i/\partial x_j) - Y^j(\partial X_i/\partial x^j)\right\}. \quad \text{[Eqn. 6].}$$

{Again, the expression XY means that X is an operator on Y where

$X = X^j(\partial / \partial x^j) = X^j \partial_j$. X^j are contravariant components and $\partial / \partial x^j$ is the coordinate <u>basis</u> of the vector X}.

Note, of course, that Lie derivative [X,Y] bears no resemblance to a **covariant derivative**:

$$\nabla_v Y = v^j e_i (\partial y^i / \partial x^j) + v^j y^i \Gamma^k_{ij} e_k,$$

where e_k is a basis like $\partial / \partial x^k$ and Γ is a Christoffel "connection" – a correction term for how basis vectors rotate under translation (like in a curvilinear coordinate system).] The Lie derivative and Lie Bracket are "independent of any particular choice of connection." [Penrose].

That is, Lie derivatives don't use the connection at all. They operate on the notion of evaluating a vector field along an integral curve of another vector field, this is inherently different to the notion of parallel transport [stackexchange].
"Look at what happens when you take the commutator of integral curves, you get the Lie derivative. On the other hand if you take the commutator of parallel transport, you get the curvature tensor."

The Lie derivative depends, not only on the value of the vector at the point p, but also depends on the value of the vector in the neighborhood of p, and thus is not a conventional directional derivative {Quora}. A problem with simple directional derivatives on a curved manifold is that the resulting vector $\nabla_W V$ will often not "lie in any tangent space to the manifold." But Lie derivatives using [W,V] will always lie in the tangent space because the off-surface components cancel out. The alternative "covariant derivative" using connections will also always lie in the tangent space.

<u>Examples of Calculating the Lie derivative $\mathcal{L}_X Y_{(p)} = [X,Y]_{(p)}$</u>

Ex. 6 <u>Easy Example:</u> Suppose we are given vector field $X = (y \partial_x - x \partial_y)$ and $Y = x^2 \partial_x$: Simply plug into the commutator [X,Y] = (X)(Y) - (Y)(X). Some terms will cancel out leaving just the answer $[X,Y] = 2xy \partial_x + x^2 \partial_y$ -- another contra-variant vector.
{<u>Note</u> that we treat $Y = x^2 \partial_x$ as a <u>product</u> of x^2 times a base ∂_x,
so $y \partial_x Y = yx^2 \partial_x \partial_x + 2xy \partial_x$, and all the double ∂'s will always cancel out}.

<u>Ex. 7 Example: Find the Lie derivative</u> for $X = (x \partial_y - y \partial_x)$

[Benn,p153] $= + \partial_\theta$ in polar coordinates. Let another field $\mathbf{Y} = (x^2 \partial_x + xy \partial_y)$. p = (x,y).
{This is the hardest but most illustrative problem so far}.
This X field is like a "pin-wheel" rotated in the CW direction say by little rockets (tangent vectors). Note that this <u>simple case</u> of a rotating point at fixed radius is $d\phi/dt = X(\phi(t))$ with $X = + \partial_\theta$. So $d\phi/dt = +d\phi/d\theta$ implies t = θ+c, and we can set c = 0. So t or time is essentially an angle θ. The easy commutator calculation $[X,Y] = - xy \partial_x - y^2 \partial_y$ is seen to be another vector.

The base vectors will be altered by the flow field: $\partial / \partial x^\alpha \rightarrow \partial / \partial x^\alpha + \mathbf{X}^\mu{}_{,\alpha}\, \partial / \partial x^\mu$ [Burke, p124]. Applied here gives $\partial_x \rightarrow \partial_x + 1 \partial_y$ (and also $\partial_y \rightarrow \partial_y - 1 \partial_x$). This is intuitively obvious for a ∂_θ circular flow of water adding another y direction vector to x and another –x direction vector to y. This is a flow change from field **X** rather than a transformation from mapping function F.

BUT, we also want to do it the other way too --- **using limits and integral curves:**
$\mathcal{L}_X \mathbf{Y} = \lim_{t \to o} [\phi_{-t*} Y_{\phi x} - Y_x]/t$ [See the arrows shown in Figure 1 above].

We first calculate integral curves of the X flow with the streamline mapping function ϕ_t . The Y_ϕ evaluates Y at a future point in time t, and the ϕ_{-t*} is a tangent map (sometimes called $d\phi^{-1}$) pushing the tangent vector back to its original point, p, for comparison.

The field X here happens to be minus that of our previous "Case 3" above. For a starting point at $\phi_t(0) = p = (a,b) = (a\,\partial_x, b\,\partial_y)$, then

$\phi_t = (\phi^1{}_t, \phi^2{}_t) = (a \cos t - b \sin t,\ b \cos t + a \sin t)$; __ **"Equation 7_$\phi_t$."**

{this is the correct equation for positive time flow, t > 0}

This is an example of X generating a 1-parameter affine group and is just a rotation matrix for rotation of axes by angle –t.

If ϕ_t of point $p = (p_x, p_y) \to q = (q_x, q_y)$, $d\phi_t/dt = (-a \sin t - b \cos t, -b \sin t + a \cos t)$ which is seen to be $(-q_y, +q_x)$ or {just like **X** itself = $(-y, x) \cdot (\partial x, \partial y)$. }

In terms of coordinates (∂_x, ∂_y): (a,b) for $Y = (x^2, xy)$. Plug these a,b's into Eqn. 7_ϕ_t to get moved ahead coordinates (a', b') for $Y_{\phi x}$. **Differentiate d/dt(ϕ_t) to get the tangent push ahead** (or back) ϕ_{-t*} using –t for the inverse. [Benn, p153, 165].
If we have successive applications of small time flows, then the resulting (x',y') from applying eqn 7 is input into eqn 7 again (x',y')→ (a', b') –e.g., a'= old (a cos t – b sin t). If the next flow is –t, then the next application will be (x",y") = the original (x,y) [using $\sin^2 + \cos^2 = 1$].
So the new (a',b') go into this new equation with the result $\phi_{-t*} Y_{\phi x} = (-xy, x^2)$ (again using $\cos^2 t + \sin^2 t = 1$). This might have been expected since the driving vector field X = (y, -x) -- reversed from the coordinates of p = (x,y).

Proceeding through the details of calculating the Lie derivative by limits:

A first step is to use the useful coordinate formula given previously:

$F_* X_p = X^i (\partial / \partial x^i) F^j (\partial / \partial \eta^j)|_{q = F(p)}$. [Benn, p 148].

This looks like the push forward begins with the operator X, but the X^i could be placed at the end. What matters formally is the usual Jacobian $[\partial \phi^j / \partial x^i]$ for push forward and $\phi = (\phi^1, \phi^2)$. We insert $Y^1 = x^2 \partial_x$ and $Y^2 = xy\,\partial_y$ in place of the X^i 's. Since our spaces are Euclidean, the new basis of N is still the old basis of M, that is $(\partial / \partial \eta)$ is just $(\partial / \partial x) = \partial_x$.

So, step one is writing the Jacobian and realizing that it no longer needs to be explicit: [Benn p 153].

$\varphi_* Y_p = x^2(p)\{(\partial_x \varphi^1)(p)\,\partial_x + (\partial_x \varphi^2)(p)\,\partial_y\}|_{\varphi(p)} + x_p y_p \{(\partial_y \varphi^1)(p)\,\partial_x + (\partial_y \varphi^2)(p)\,\partial_y\}|_{\varphi(p)}$.

$\mathcal{L}_X \mathbf{Y} = \lim t^{-1}\{(\cos t - 1)[x^2 \partial_x + xy\,\partial_y] - \sin[xy\,\partial_x + y^2 \partial_y]\}$. {the y^2 term is the tricky part}.
Answer is $\mathcal{L}_X \mathbf{Y} = -xy\ \partial_x - y^2\ \partial_y$. [i.e., (sin t) /t → 1).
[and the commutator [X,Y] is a much easier calculation].

A different elaborate step-by-step approach is developed in a text by Burke [Burke p. 164]. This selects a simplest reference point p = (0 , 0) in R2 and examines small x, y values near that zero (i.e., effectively first-order approximations). This is a derivation of the commutator from limits of flows: $\mathcal{L}_X \mathbf{Y}$ = [XY-YX].

When applied to the current problem above (tiny-circles about the origin), it yields the same Lie bracket answer as before.

Expressions for ($\mathcal{L}_X Y$)

($\mathcal{L}_X Y)_p$ = $\lim_{t\to o} [Y_p - (\phi_{t*} Y)_p]/t = \lim_{t\to o} [(\phi_{-t*}Y)_p - Y_p] /t$
{An important nuance or correction to the above: **($\mathcal{L}_X Y)_p$ =** $\lim_{t\to o} [Y_p - (\phi_{t*} (Y(\phi_{-t}(p)))]/t$ }
about the points where Y is evaluated [Felice, p 63].

$[X,Y]_x = (d/dt)((D_x\psi_t)^{-1} Y_{\psi(x)}) |_{t=0}$ *or* $\mathcal{L}_X Y_p = (d/dt)|_{t=0} (\varphi_{-t*})Y_{\varphi t(p)}$
($\mathcal{L}_X Y)_p$ $= (d/dt)(\phi_t^* Y)|_{t=0}$ (using Pullback) $= -(d/dt)\phi_{t*}Y(x)|_{p,t=0}$

Example: Lie derivative of a function, f (again) [Felice].
($\mathcal{L}_X f)_p$ $= (d/dt)(\phi_t^* f)|_{t=0} = -(d/dt)\phi_{t*}f(x)|_{p,t=0}$=$-d/dt\, [f\circ \phi_t^{-1} = f\circ \phi_{-t} = f(\phi_{-t})]$
$= +d/dt(f(\gamma_p(-t)))= d\gamma/dt\, f = X(p)f = X(f)_p$. Which is again just the directional derivative.

<u>Derived Properties of the Lie Derivative:</u>
$\mathcal{L}_X f = Xf$, $\mathcal{L}_X (\phi) = X(\phi) = X^\alpha \partial_\alpha \phi$.
$\mathcal{L}_X (\partial / \partial x) = -\mathcal{L}_{(\partial / \partial x)} X = (\partial / \partial y)$?? or minus -?
$\mathcal{L}_{(\partial / \partial xi)} (\partial / \partial x^j) = [(\partial / \partial x^i), (\partial / \partial x^j)] = 0$
$[\mathcal{L}_X , \mathcal{L}_Y] = \mathcal{L}_{[X,Y]}$.
$\mathcal{L}_X (fg) = (\mathcal{L}_X f)g +f \mathcal{L}_X g$, and $\mathcal{L}_X (fY) = (\mathcal{L}_X f)Y+f \mathcal{L}_X Y$.
$\mathcal{L}_{X+Y} = \mathcal{L}_X + \mathcal{L}_Y$.
$[\mathcal{L}_X , d]=0$, $\mathcal{L}_X d = d \mathcal{L}_X$ (d=exterior derivative).
$\mathcal{L}_X \omega = i_X d\omega + d(i_X\omega)$ {Cartan formula for exterior forms –
see "contractions," i_X, in the Definitions section}.
$d\mathcal{L}_X \omega = \mathcal{L}_X (d\omega)$. {one form fields}

<u>The Lie derivative also applies to a 1-form field</u>

The concept is: "Lie-derivative = Lim(value pulled back - value already there" [Burke]. For a 1-parameter transformation $\Phi\epsilon : M \to M$; point $q\to q_\epsilon$; the Lie derivative of a 1-form ω is $= \mathrm{Lim}_{\epsilon\to o} (\Phi^* \cdot \omega(q_\epsilon) - \omega(q))/\epsilon$ with an asterisk <u>superscript</u> for pull-backs. This resembles using V* for dual vectors of V. In general, a mapping ϕ^* takes p-forms on **N** to p-forms on M (the opposite direction from the ϕ_* mapping for tangent vectors); but here we are only considering p = 1 or 0 forms.

A 0-form is just a function, $f = f(y^\beta)$. $f^* = f\circ\phi$ or $f^*(x^\alpha)=f(y^\beta(x^\alpha))$.
[Gibbons] $\partial f^*/ \partial x^\alpha = (\partial y^\beta/ \partial x^\alpha) \partial f/ \partial y^\beta$, $\phi^*\omega = {}^*\omega$, a moved object.
${}^*\omega_\alpha(p) = (\partial y^\beta/ \partial x^\alpha)\, \omega_\beta(p')$ -- using the Jacobian, J, transformation again.
An exterior derivative will commute with pullback: $d(\phi^*\omega) = \phi^*(d\omega)$.

<u>Ex. 8 Example</u> <u>of Lie derivative calculation for one-forms:</u> [math.stackexchange]
Let $X = a\, \partial_\theta$ and ω = 1-form $\sin\theta\, d\theta\wedge d\phi$

and mapping $\psi_{t,\, a\partial\theta}(\theta,\phi) = (\theta + at, \phi)$,
$\mathcal{L}_X \omega = (d/dt)|_{t=0}$ **{** $\psi_{t,\, a\partial\theta}$ *(sinθ dθ∧dϕ) =sin(θ+at)dθ∧dϕ }=
a cos(θ +at)$|_{t=0}$ dθ∧dϕ = a cos θdθ∧dϕ. (answer).

Some Definitions and Background Material:

Notation: There is much variation and little standard convention in symbols used in Differential Geometry. But here is a possible sample:

Curves: α, β , γ	Vector fields: V, X, Y, W.
Mappings: F, G, (others use ϕ, ψ)	Functions: f, g
1-Forms: ψ , φ , ω	2-Forms η , φ∧ ψ = φ ψ
Points p, q, x_o	Bases: $dx_i(U_j) = \delta_{ij}$.
M Coordinates F: M→ N, x^j	Coordinates in N: y, η, $\bar{x}$ $\bar{A}$, $\bar{y}$

Vector V = $\Sigma v^i U_i = \Sigma v^i \partial_i = \Sigma v^i\ \partial/\partial x^i$.

1-1 correspondence on R^3: $\Sigma f_i dx_i \leftrightarrow \Sigma f_i U_i \leftrightarrow f_3 dx_1 dx_2 + f_2 dx_3 dx_1 + f_1 dx_2 dx_3$. {star,∗,⋆}

A **curve** is a differentiable or "smooth" function γ : I →M from an open interval into a manifold M. Parameterizations of curves might select an interval [0,1] or [0, 2π] on the real line $\mathfrak{R}^1$, but any real interval is allowed. For example, the helix curve might have a parameter t ∈ $\mathfrak{R}$ over all reals: γ(t) = (x(t), y(t), z(t)) = (a cos t, a sin t, bt) .

A speed of γ(t) is γ'(t) = ($d\gamma_1(t)/dt$, $d\gamma_2(t)/dt$, $d\gamma_3(t)/dt$) or (d/dt)(x(t), y(t), z(t)). A tangent vector is the vector speed of some curve for some parameter (e.g., t for time). Unlike elementary Euclidean 'vectors,' it has two parts: its point of application p and its vector part v. Other names for tangent vector are just "tangent" and "contravariant vector." The space of tangent vectors at point p is called $\mathbf{T}_p$.

Covariant versus Contravariant: definition .. any set of quantities transforming according to the following form:
$\bar{y}^i = (\partial \bar{x}^i/\partial x^k)x^k$ =Jacobian· x^k is called contravariant. $\bar{x}^i$ is a function of the x^k's [Adler, GRT 1965]. Other notation may be: $v^i = (dx^i/dt) = (\partial x^i/\partial x^k)(dx^k/dt)$. Transforming a function f(x^i's): $df = (\partial f/\partial x^i)dx^i$ … is contravariant. {sometimes called a "famous classical formula"}.
Covariant goes like $\bar{A}_i = (\partial x^k/\partial \bar{x}^i)\, A_k$, and 1-forms obey $\omega'_i = (\partial x^k/\partial x^i)\, \omega_k$. {So, are the last two indices the same ? [contravariant] or different [covariant]. Contravariant components have indices high and covariant components have indices low.

But the **bases** $\partial_x = \partial/\partial x$ for contravariant vectors transform like the covariant form: $\partial_k = (\partial x^i/\partial x^k)\,\partial_i$ -- so we might write the push-forward of a base as
$\psi_* : \partial_x \rightarrow (\partial Y^\alpha/\partial x)(\partial/\partial y^\alpha)$ [Burke p 79]. And a form-base pull-back may look like:
ψ* dy → $\partial Y/\partial x^\alpha$) dx^α (which looks contravariant).

A vector field **Y** on a curve γ : Interval I →M is a function that assigns to each number t ∈I a tangent vector Y(t) to M at the point p = γ(t). Y(t) = (y_1(t), y_2(t), y_3(t))$_{\gamma(t)}$.

Mathematicians like to emphasize generality (e.g., the term "wlog" means 'without loss of generality') and work from most primitive structures upwards. We all learned the definition of continuity as:

$\forall$ $\epsilon > 0$, $\exists$ $\delta > 0$ such that metric measure $d(x,y) < \delta \Rightarrow d(f(x),f(y)) < \epsilon$.

But that depends on having a "proximity function" such as |x-y|. Topology uses a more primitive approach based on open sets.

For function f: U→W and continuity at point p: { let "nbh" be an open neighborhood}

$\forall$ nbh W' with f(p)$\in$ W', $\exists$nbh U' containing f(p) whose image f(U')$\subseteq$ W'.

{"nbh" = neighborhood [Benn] }. "A space with topology defined on it is called a topological space." "If a map between topological spaces is continuous with a continuous inverse, it is called a homeomorphism." Then we add the Hausdorff property that: "disjoint neighborhoods can be defined about distinct elements of the space. We also need the space to be locally homeomorphic to an open set of $\Re^n$ ("locally Euclidean"). Then we have a "topological manifold."

A manifold is a set of points with "neighborhoods." A differentiable manifold has the property of being locally similar enough to a linear space to allow one to do calculus [Wik]. One can have a coordinate system for the neighborhood (but it may take multiple charts to cover the whole manifold).

A **curve** α(t) "in R^3 is a differentiable function $\alpha : I \rightarrow R^3$ from an open interval I into R^3" (O'Neill). For the special case of relativity, we re-parameterize h: $J \rightarrow I$, t→ α (t) where the parameter in J is arc-length s = tau = τ (proper-time). For simple lines, we might have α(t)= vt , and (since dt/dτ = gamma, γ) t = γ τ . Then velocity α'=dα/dτ = (dα/dt)(dt/dτ) = γα'. The 4-velocity is U = dx/dτ = γ (c,**v**). {And, "a material particle α in spacetime M **is** a future-pointing timelike curve α: $I \rightarrow$ M." [ONeill-Kerr] }

Mappings: If F is a mapping from manifold M to N and v is a tangent vector to M at point p. $F_*(v)$ is the initial velocity of the curve t→ F(p+tv). A **tangent map** takes a tangent vector in M and maps it to its corresponding tangent vector in N. [ONeill].

{ Math expression) For mapping ϕ: M→N, $\exists$ $d\phi : T_p(M) \rightarrow T_p(N)$. The differential map dϕ preserves tangents: for a curve $\alpha \subset M$, "dϕ carries each vector α' (s) to the tangent vector $(\phi \circ \alpha)'$ (s) of the image curve $(\phi \circ \alpha) \subset N$ [ONeill_Kerr]." }

The "Jacobian" J = J_{ij} :

J is a generalization of the gradient concept which became the matrix of all first-order partial derivatives of a vector-valued function, also Df, J_f, Few sources say what the Jacobian matrix operates on – but for several cases it is column vectors.

The Jacobian of a scalar function is the transpose of its gradient (that is, ∇f = a column vector but J = [row vector].)

For mapping F: manifold M → N, if mapping F = $(F^1(x,y), F^2(x,y))$, then the first row of J is conventionally $[\partial F^1/\partial x, \ \partial F^2/\partial x]$. In general for "push aheads" it is $(\partial F^i/dt) = [\partial F^i/\partial x^j](\partial x^j/\partial t)$ where $[\partial F^i/\partial x^j] = J_{ij}$. We say that mapping F at point p induces a tangent vector mapping F_{*p} such that $(F_{*p} V)(f) = V(f \circ F)$. Or, tangent velocity $V_{\text{on } N} = J V_{\text{on } M}. = F_{*p} (V)$.

Integral Curves and "FLOW:" (motivated by time-independent flow of water in $\Re^3$ telling how the individual water molecules are transported) [Frankel]. Each vector field has an associated flow $\{\phi_t\}$ having v as its velocity field. The flow describes the "integral curves" or streamlines of transport points from one time to another; and these are solutions of the differential equation:

$\partial\ \boldsymbol{\phi_t(p)}/\partial\ \mathbf{t = v(\phi_t(p))}$ **<u>or</u>** $\mathbf{dx^i(t)/dt = V^i(x^i)}$

{--called "The fundamental theorem on vector fields"}.
$\phi_t(p)$ moves a point p to a later point. An intention is to restrict the range of parameters so that we can talk about a "local flow" – a 1-parameter group of diffeomorphisms so that $\phi_t^{-1} = \phi_{-t}$. For short parameter ranges t and s, $\phi_t \circ \phi_s = \phi_{t+s} = \phi_s \circ \phi_t$.

"In mathematics, an integral curve is a parametric curve that represents a specific solution to an ordinary differential equation or system of equations. If the differential equation is represented as a vector field or slope field, then the corresponding integral curves are tangent to the field at each point." [Wik]

"<u>Derivation</u>" as a term-- is a generalization of the derivative operator. A derivation is an operator on an algebraic system which is linear and obeys the product rule (Leibnizian: V(af+bg) = aVf + bVg). This includes the exterior product: **d** is a derivation. i_X contraction or interior derivative, partial ∂ is an R-derivation; L_X Lie derivative; and a linear map on p-forms $\Lambda^p M^n \rightarrow \Lambda^{p+r}$ if r is even (else an anti-derivation with a negative term). "A vector field X on a manifold M is a derivation on the algebra of smooth functions" [Benn, 142].

In differential algebra, a <u>derivation</u> is just the linear differential operator term of the Taylor's series expansion of a mapping and obeys the Leibniz rule for derivatives: D(fg) = fD(g) + gD(f). In differential geometry, derivations are tangent vectors. The differential dx represents an infinitely small change in the variable x. The differential of a mapping at a point p is a linear approximation of the mapping near p – it is sometimes called a "push forward."
Df(p) is a linear transformation, so F_{*p} is a linear map on tangent space T_p. Some conventions place the asterisk lower, f_* or φ_*. If a mapping has a smooth inverse, it is a diffeomorphism.

FORMS: *Forms are intuitively described in a variety of ways:*

a) They could be considered as "a thing" under an integral sign: $\int$ 2xdx has a 1-form ψ = 2xdx, $\int$ 3xydxdy has 2-form $\alpha = 3xydx \wedge dy$ {that is, the 2-form has to be anti-symmetric so that dxdy = -dydx, order is important; and the "wedge" sign $\wedge$ is a reminder of that}. Of course this also implies that we can integrate forms: $\int \psi$. Note that repeats are zero, dxdx = - dxdx = 0.

b) Some view 1-forms as a "family of flat, equally spaced surfaces" [MTW], and the number $\psi v = \langle \psi, v\rangle$ is the "number of surfaces pierced" by a vector v passing through the surfaces ψ {the "bongs of a bell" – one for each piercing}. For de Broglie waves, a 1-form κ (made from wave-number k) may be made of surfaces of constant phase (on a sine function). Then $\langle k,v\rangle$ is a phase difference. κ is the gradient of a function for advancing phase, $\kappa = d\phi$. κ is a "machine into which vectors are inserted and from which numbers emerge." A directional derivative is $\partial_v f = \langle df, v\rangle = v_p[f]$ with differential df.

c) In differential geometry, "a 1-form ψ on $\mathfrak{R}^3$ is a real valued function on the set of all tangent vectors to $\mathfrak{R}^3$ such that ψ is linear at each point, that is
$\psi(av + bw)=a\psi(v)+b\psi(w)$" [O'Neill]. At a point p, ψ_p is an element of the dual space of $T_p(\mathfrak{R}^3)$ – the space of all tangent vectors.

d) The **<u>Dual space V*</u>** is the set of all linear maps ϕ: V→F (field) where ϕ is also called a form or convector. $\phi(x) = \langle \phi, x\rangle$. For example let a non-orthogonal basis of R^2 be $e^1 = (\frac{1}{2}, \frac{1}{2})$ and $e^2 = (0,1)$. Then the dual basis is $e_1 = (2,0)$ and $e_2 = (-1,1)$ or $e_1(x,y) = 2x$, $e_2(x,y) = -x+y$. So $e_1e^1 = (2x \frac{1}{2}, 0 \cdot \frac{1}{2}) = (1,0)$ [WIK]. For quantum mechanics, bra's are linear functionals on ket's: ⟨ bra,ket⟩ = a positive real number.

<u>e)</u> **<u>Cotangent Space</u>** [Felice]: "The set of all linear maps from $T_p(M)$ into $\Re$ is called the cotangent space at p, $T^*_p(M)$. The differential of a function $\omega = \mathbf{d}f_p$ is an example. $\mathbf{d}x^i$ is a basis so that any "covector" $\boldsymbol{\omega} = w_i\, dx^i$ {In very simple terms, a covector is any differential placed inside an integral sign.}

For a mapping from M to N, $\phi_* : T_p(M) \mapsto T_{\phi(p)}(N)$, and $\phi^*: T^*_{\phi(p)}(N) \mapsto T^*p(M)$ – backwards. Tensors can be decomposed into a sum of tensor products of vectors and 1-forms.

So tangent pull-backs are intended for covectors or 1-forms, $(\phi^*\omega)_p = \phi^*(\omega(\phi(p)) = \phi^*(\omega \circ \varphi)$. For vectors V, $\phi^*(\omega)(V) = \omega(\phi_*(V))$ taking covectors on N into covectors at M. So, pullbacks and pushforwards are inverses of each other, $(F^*)_* = F$. $\phi^*\omega$ is always a well defined covector field, but $\phi_* V$ is indefinite unless ϕ is 1:1.

Pullbacks are **<u>defined</u>** in terms of "PushForwards"

$\mathbf{F^*\omega\ (V)= \omega(F_*\ V) = \omega\ (dF(V)) =F^*(\omega \circ F).}$

Although this is a "definition" it may be "derived" via the common composition expression.

<u>Ex. 9 Example: Pullback φ *ω of a form ω:</u> [stackexchange]

Suppose mapping $\varphi : (u,v)\in M \mapsto (x,y)\in N$, $R^2 \to R^2$.

Lets suppose: $\varphi(u,v) = (x,y) = (uv, u^2)$, and 1-form $\omega = xy\, dx + 2x\, dy$ on N.

In terms of u and v, $dx = (\partial x/\partial u)du + (\partial x/\partial v)dv = vdu+udv$, and $dy = 2udu$,

Simply substitute these into ω(x,y) to get

$\varphi^*\omega = \omega\circ\varphi = \omega(u,v) = (uv)(u^2)(vdu+udv)\ ,\ 2(uv)(2udu) = (u^3v^2+4u^2v)du+u^4vdv.$ ⟵

So, the form ω is now expressed in variables of M pulled back from N.

Now try the "other" definition in terms of "push-forwards": $\phi^*\omega = \omega(d\phi)$ – note "a <u>function of</u>" rather than product ω dϕ .

For $d\phi = (d\phi_1, d\phi_2) = (dx, dy) = (d(uv), d(u^2)) = (udv+vdu, 2udu)$.

$\omega = xy\, dx + 2x\, dy \to uv\, u^2 (udv+vdu) + 2(uv)(2udu)$ = (same answer as above). ⟵

Replace all the x, y, dx, dy of variables for N with u, v, du, dv variables of M.

If ϕ: M→ N, and f: N→ $I\subset \Re$, then the pullback of this smooth function f is just $(\phi^*f)(x) = f(\phi(x)) = (f\circ\varphi)(x)$. f can be considered a 0-form.

Use of "**<u>Contractions</u>**" <u>i_v</u> **<u>on forms</u>** (i for "inner product"): [Frankel]

The notation α^p stands for a p-form. If α is a covariant vector (α^1, a 1-form) and v is a contra-variant vector, then $\alpha(v) = \alpha_i v^i$ is a scalar (a 0-form, and α^o is a 0-form). {Recall that for a tensor T that is p times contra- and q times co-, a contraction $T^{\ldots i\ldots}{}_{\ldots i\ldots}$ is (p-1) times contra and (q-1) times covariant – <u>a reduction in rank</u>}

$i_v\alpha^o = 0$, $i_v\alpha^1 = \alpha(v) = \alpha_i v^i$ (like a dot product).

Contraction of a function: $i_X df = X(f) = \mathcal{L}_X(f)$.

A volume element in $\Re^3$ can be a 3-form (n-form), $vol^n = \rho(u)du^1 \wedge \ldots \wedge du^n$ [Frankel p.90,120]. Its contraction with a vector is a special object called a "pseudo-2-form" like the magnetic field β (recall that B = curl A is not a vector but rather a **pseudo**-vector that changes sign under mirror reflection).
$\beta^2 = i_B\, vol^3 = B_{23}\, dx^2 \wedge dx^3 + B_{31}\, dx^3 \wedge dx^1 + B_{12}\, dx^1 \wedge dx^2$. " 2 is 3 -1 form".
"The following is perhaps the most often used formula involving Lie derivatives" when acting on exterior forms—Cartan's Formula: $\mathcal{L}_X = i_X \circ d + d \circ i_X$.
Then [Wik] $\mathcal{L}_X f = i_X df$, $\mathcal{L}_X \omega = i_X d\omega + d(i_X\omega)$, $d\mathcal{L}_X \omega = \mathcal{L}_X (d\omega)$.

Ex. 10 Example interior product of vector field X on 2-form dω [stackexchange].
Let $\mathbf{X} = y\,\partial_x + 2z\,\partial_y + 3xy\,\partial_z$. And 1-form $\omega = 3xdy - 7zx^2dz$, so
dω $= 3dx \wedge dy - 14zx\, dx \wedge dz$ (using $dz \wedge dz = 0$).
Just multiply Xdω through using $\partial_x dx = 1$, $\partial_i dx^j = \delta_{ij}$.
Or, more conveniently,
$i_X d\omega = X^j (d\omega)_{ji}\, dx^i = X^1 d\omega_{12} dx^2 + X^1 d\omega_{13} dx^3 + X^2 d\omega\, (-)_{21} dx^1 + X^3 d\omega (-)_{31} dx^1$
$= (y)(3)dy + (y)(-14zx)dz + (2z)(-3)dx + (3xy)(--14zx)dx$
$i_X d\omega = 3ydy - 14xyzdz - 6zdx + 3 \cdot 14x^2 yz\, dx$ -- a one form (p-1=2-1=1).

Comment: A very concise and modern summary of the above is given in Barrett O'Neill's book on Kerr Geometry [ONeill_Kerr] Chapter one on general background material.

Applications of Lie Derivatives:

Practical applications of the Lie Derivative seem to be more limited than Covariant derivatives (e.g., General Relativity).

1. Fluid Flow, [Frankel, 143] Velocity co-vector $V = v_i dx^i$, momentum density ρv, Momentum $P = \int v_i \rho\, vol^3$. $\mathbf{X} = (\mathbf{v} + \partial / \partial t)$ vector field. $\mathcal{L}_X (\rho vol^3) = 0$. Total force is $dP/dt = \int [\partial v^i / \partial t + v^j (\partial v_i / \partial x^j)] \rho vol^3 = \int X(v_i) \rho vol^3 = \int \mathcal{L}_X (v_i \rho vol^3)$. So, since each velocity component is just a function, we can express force in terms of $\mathcal{L}_X$.
2. "Advection" is material transport via bulk motion. Let **u** be a fluid velocity vector field and ψ a relevant scalar quantity in its flow. There is a continuity equation,

$\partial \psi / \partial t + \nabla \cdot (\psi \mathbf{u}) = 0$ and an "advection operator" $\mathbf{u} \cdot \nabla$. The "Material derivative" is defined as $D/Dt = \partial / \partial t + (u \cdot \nabla)$ "and it computes the rate of change of say a time dependent vector field along the flow as $D\mathbf{u}/Dt(\phi_t(p)) = \partial\, \mathbf{u}\, (\phi_t(p),t)\, \partial t$." That is, $u \propto dt$ "on any field line." Material vector field **u** is an invariant field with its lines "frozen into the fluid." [Childress] For steady flows, $L_u v = (d/dt)(\phi_t^* v)|_{t=0}$ in terms of the "pullback." The Lie bracket can be calculated from this [pg. 5]. $\mathbf{v}(x,t)$ is frozen into a fluid if $v_t + [u,v] = 0$.

For 1-forms, $\partial \omega / \partial t + L_X \omega = 0$.

"the spatial Lie derivative is an underlying element in all areas of **mechanics**: for example, the rate of strain tensor in elasticity and the vorticity advection equation in fluid dynamics are both nicely described using Lie derivatives." [ArXiv 0912.1177 Mullen]
[Boyland] The gradient vector field $\nabla\alpha$ is defined using the metric as the unique vector field that satisfies $d\alpha(v_p) = g(\nabla\alpha, v_p)$ for all vectors v_p.

In differential topology, “The importance of the commutator form [,] for Lie derivatives comes from Frobenius' Theorem, which tells us when a given distribution of vector fields in a manifold can be "integrated" to form the tangent bundle of a submanifold. This is a really, really, really important theorem. But, depending on the text you're reading, that importance is not always evident.” {i.e., For an open set U and F a smooth differentiable 1-form on U, “the Frobenius theorem states that F is integrable if and only if for every p in U the stalk F_p *(of a sheave)* is generated by r exact differential forms [Wik] }.

Killing Vectors:

Killing vectors are named for a Norwegian mathematician named W. Killing, who first described these notions in 1892. They are vector fields that preserve the metric, g, and so are infinitesimal generators of isometries. Distances of objects are preserved along Killing vectors. If X is a Killing vector, then $\mathcal{L}_X g = 0$.

An easy example [Wik] is the upper-half x,y plane “M” with the metric: $g = dr^2/y^2 = (dx^2+dy^2)/y^2$ where (M,g) is called the hyperbolic plane and has Killing vector field ∂_x. The metric in $\mathfrak{R}^2$ is diagonal with only g_{11} and g_{22} components.

$\mathcal{L}_X g = 0$ so that distances are unchanged along Δx displacements.

That is: $\mathcal{L}_V g = V^a \partial_a g_{\mu\nu} + (\partial_\mu V^a) g_{a\nu} + (\partial_\nu V^a) g_{\mu a} = \nabla_\mu V^\nu + \nabla_\nu V^\mu$.

But $K = \partial x$ with $\mathcal{L}_x g = 0$.

If there is a “tetrad” formalism, then Killing vectors can be computed using the Cartan homotopy formula “id + di”).

Metrics in mathematics are supposed to be positive measures. But in relativity, a metric g can have negative values -- as in “time-like” rays for $ds^2 = d\sigma^2 - dt^2$. Such metrics are called semi-Riemannian or “pseudo-Riemannian.”
The generalization of a straight line in Euclidean space is a “geodesic” on curved spaces or on semi-Riemannian manifolds.
“A curve γ in M is a geodesic provided its acceleration is zero: γ ‘’ = 0 ” where ‘prime’ denotes derivative with respect to the parameter s of curve γ(s).
The geodesic equation is: $x^{k''} + \sum \Gamma^k_{ij} x^{i'} x^{j'} = 0$.

“If X is a Killing vector field on M and γ is a geodesic, then the scalar product $\langle X, \gamma' \rangle = g(X, \gamma')$ is **constant** along γ.” [ONeill-Kerr p 20].

A Killing vector leaves the metric unchanged under infinitesimal coordinate changes (e.g., from t→ t+dt). The Schwarzschild metric $g_{\mu\mu}$ in general relativity has no dependence on variables t or ϕ so that examples of its Killing vectors can be $K_1 = \partial_t = (\partial / \partial t)_{r,\theta,\phi \text{ (held constant)}}$ and $K_2 = \partial_\phi = (\partial / \partial \phi)_{t,r,\theta}$ Or, sometimes, $K_1 = (g_{tt},0,0,0)$ and $K_2 = (0,0,0,g_{\phi\phi})$. These are also two of the Killing vectors for the rotating Kerr spacetime in Kerr coordinates.

The contravariant Killing vectors are just $K^\mu = \delta^\mu{}_\nu$ just ones and zeros, $K^1 = (1,0,0,0)$ and $K^2 = (0,0,0,1)$. $K_\mu = g_{\mu\nu} K^\nu$. In another note [MTW], $K_t \cdot K_t = g_{tt}$ and $K_\phi \cdot K_\phi = g_{\phi\phi}$ (for Kerr-Newman black holes).
{Another source, Matt Visser, The Kerr Spacetime: A brief introduction, https://arxiv.org/pdf/0706.0622.pdf gives the same results (1,0,0,0) and (0,0,0,1)}.

Killing vectors can be computed in tetrad formalism easily if one uses the Cartan homotopy formula, $L_V=i_V d + d i_V$ (id+di) {Contractions, but we won't discuss tetrads here}.

References:

[Frankel] Theodore Frankel, The Geometry of Physics, An Introduction, Second Edition, Cambridge, 1997, 2004.

[ONeill] Barrett O'Neill, Elementary Differential Geometry, Academic Press (1966, 1997, 2006).

[ONeill_Kerr] Barrett O'Neill, Geometry of Kerr Black Holes, 1995 or Reprint Edition 2014, (Dover Books on Physics) {The Amazon look-ahead on general background material should be carefully studied! Wading through this is a test of how well you understand}.

[Burke] William L. Burke, Applied Differential Geometry, Cambridge,1985

[Wik] https://en.wikipedia.org/wiki/Lie_derivative

[Penrose] Roger Penrose, The Road to Reality, a complete guide to the laws of the universe, Knopf, 2005

[Benn] I M Benn, R W Tucker, An Introduction to Spinors and Geometry with applications to physics, Adam Hilger, 1987.

[Wald] Robert M. Wald, General Relativity, Chicago, 1984

[MTW] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler, GRAVITATION, Freeman, 1970-1973

[Felice] F. De Felice and C.J.S. Clarke, Relativity on curved manifolds, Cambridge, 1990.

[Boyland] Philip Boyland, "Fluid Mechanics and Mathematical Structures," https://people.clas.ufl.edu/boyland/files/asiweb.pdf

[Childress] Stephen Childress, "Topological fluid dynamics for fluid dynamicists." https://www.math.nyu.edu/faculty/childres/tfd.pdf

[MathStackExchange] questions and answers on the web:

https://math.stackexchange.com/questions/1524410/computing-lie-derivative $Y=x^2\partial_x$.

https://math.stackexchange.com/questions/435801/calculate-the-lie-derivative, unused exp(-5t).

https://math.stackexchange.com/questions/3081246/worked-examples-of-lie-derivatives Forms!

https://math.stackexchange.com/questions/576638/how-to-calculate-the-pullback-of-a-k-form-explicitly

http://bose.res.in/~amitabha/diffgeom/chap11.pdf

http://www.damtp.cam.ac.uk/research/gr/members/gibbons/gwgPartIII_DGeometry2011-2.pdf Part III: Applications of Differential Geometry to Physics

Killing vectors: https://physics.stackexchange.com/questions/214434/killing-vectors-in-schwarzschild-metric

Ebrahim Ebrahim, "A Self-Contained Introduction to Lie Derivatives," http://web.math.ucsb.edu/~ebrahim/liederivs_tame.pdf

{ $\mathcal{L}$ Lucinda Calligraphy} and "overbar" is in insert equation, accents. $\bar{x}$ $\bar{A}$ $\bar{f}$ $\bar{y}$

Long Spin Disk Lube Migration

David Peterson, *Member, IEEE*

Abstract—**The open literature on surface lubricant depletion of magnetic data storage hard disks has generally examined nonporous media having no inner diameter hole and short spin times (e.g., a few hours to a few weeks). Applied surface lube was usually thick (above 10 nm) and subject to inertially driven migration. Here we report the depletion of thin surface lubricant (<5 nm) on commercially used particulate porous media with radial migration driven by air shear stress over a six-year spin period. The radial thickness profile of the outer disk pack assembly surface lubricant after long spin time is characterized by nearly complete depletion near the inner diameter followed by a ramp up to a semi-stable outer annulus plateau. ID depletion does not occur on interior pack corotating surfaces because of nearly solid body air rotation near the hub and the associated reduction of air shear stresses there. Elementary modeling is done for the development of the outer pack surface profile. One such model includes surface lurbricant replenishment from subsurface porosity, surface air shear stress, and effective interface slippage of the surface lubricant.**

I. Introduction

WINCHESTER rigid disk drives require disk surface lubrication largely because head read/write sliders come in contact with the disks during start/stops. Unfortunately, nonbonded liquid lubricants have often been observed to migrate away from the inner annuli of outer pack disks, thus producing disk underlubrication conditions which contribute to "head crashes." The thickness of lubricant above the surface of commercial disk's is generally very thin (from 1.5–10 nm). For such thicknesses, the primary driving force on outer-pack surfaces of head/disk assemblies (HDA's) is the "air shear stress" due to the pumping action of a spinning disk pack [1]–[4]. In this paper, we will be primarily concerned with the long term migration radial profiles of surface lubricant on outer pack disk surfaces. Inner disk pack (or "corotating") surfaces have a different migration profile near the disk ID which may be driven by a combination of weak inertial forces and weakened air shear stress. The air near the ID of inner pack disks is in nearly solid body corotation, and that means that air shear stesses are strongly diminished there. Even at the outer pack surfaces, air shear stress is somewhat attenuated from that of a spinning "free disk" due to the proximity of baseplates and plastic shell shroud. Outer pack surfaces often evolve a "ramp" profile of surface lubrication thickness which separates a depleted inner annulus from a higher "plateau" level outer annulus. Disk pack interior surfaces usually lack this profile, and the surface lubrication thickness of the entire surface is at the plateau level. Many factors affect the nature and rate of disk lube migration. These include: the lubrication capacity of coating porosity, the size distribution of pores, the relationship between surface and subsurface lubricant thickness, the disjoining pressure of the thin lubricant film [5], [6], surface hydrocarbon contamination, disk surface roughness [7], laminar to turbulent air flow transition, the resulting surface radial air shear stress profile ($\tau(R)$), coating thickness radial profile, lubricant viscosity and its modification near the interface, surface lubricant diffusion rate, possible interface slippage [8], possible lubricant evaporation [9], lubricant molecular weight distribution, and the existence of an ID boundary initial condition in addition to the initial surface lube thickness distribution profile. It is impractical to include so many factors in a precise model of disk lube migration. This is especially true since such important concepts as disjoining pressure and polymer slip are not currently well characterized. To account for observed development profiles with time, we use an "engineering model" which relies primarily on three parameters: the effective efficiency or coupling of air shear stress ($E < 1$), an effective interface lube slippage (or "extrapolation length," B), and an approximate proportional replenishment of surface lube to provide a balance with subsurface lubrication (with P = surface/total lube thickness). It is understood that these few convenient parameters may incorporate some of the effects of the much more complex underlying reality.

Manuscript received October 10, 1991; revised January 30, 1992.

The author is with Storage Technology Corporation, 2270 S. 88th Street, Louisville, CO 80028-1103.

IEEE Log Number 9107736.

II. Experimental Results

A. Surface Lubricant Thickness by ESCA

3380 type disks from head/disk assemblies were examined after spin times up to 77 mo under normal usage drive operating conditions. These HDA's contained 356 mm (14 in) diameter hard particulate media referred to as "type C2" lubricated with a "Fomblin type Y" fluorocarbon lubricant (see Appendix I for details). Selected outer pack disks were systematically analyzed by radial scan "15-point-line-ESCA" (electron spectroscopy for chemical analysis [10]) for the radial surface lubricant thickness profiles, $h(R)$. The calibration of ESCA is ensured by a standards program at SSL (Surface Science Laboratories). "ESCA thickness" is intended to measure the lubricant thickness above the binder surface, h, rather than the total lube thickness throughout the porosity, H. The ability to determine h by ESCA techniques depends

0018-9464/92$03.00 © 1992 IEEE

on knowing the escape depth of photoelectrons in the fluorocarbon lubricant. Conventional measurements (the "Tau method" [10]) assume that this value is 4.0 nm. A recent publication [11] reports an actual measurement at 2.9 nm. This means that all conventional reported values of "ESCA lube" may have to be scaled down by a correction factor, $f \approx 1.4$. We will continue using the symbol h for conventional ESCA surface lubricant measured value thickness but will use "real h" $\approx h/f$ inside computer programs modeling theoretical migration.

A typical set of long spin surface lube thickness profiles is shown in Fig. 1. Profile "0" near $h = 4$ nm average surface lube thickness represents the initial state of the lubricant at zero spin time. This profile is an average of ten sample surfaces processed by "point-line" radial sector ESCA and the older individual point "punched sample" ESCA analyses. Curves "A" and "B" show typical ramp profiles for outer pack surface lube migration after 77 mo of HDA spin time. Nearly complete depletion of lubrication has occurred at the inner annuli of the disks to a disk radius referred to as "$R1$." This is followed by a ramp up to an outer annulus plateau level of higher thickness surface lubrication beginning at a radius called "$R2$." The point where the ramp joins the plateau is called the "vertex." The outer plateau may also represent the decrease from an initially uniform $h(R)$ to semiuniform lube levels that would have resulted from spinning disks which had no inner diameter hole. The difference in depletion radii ($R1$'s) for curves A and B may be due to differences in surface roughness of these two outer-pack surfaces (roughness average $R_a[A]$ as measured by mechanical profilometry was $\sim 19\%$ rougher than was $R_a[B]$). Curve "C" is an average of the two opposite side interior pack surfaces and shows typical lack of surface inner annulus depletion.

Fig. 2 is a detailed plot showing another typical ramp from a depleted ID annulus which has just formed on an outer pack surface at 14.5 mo of HDA spin time. The extent of this ramp is 12 mm—a typical value for initial formation ramps. The dashed curve of Fig. 2 shows that lube depletion has not occurred on the opposite "interior surface" side of the disk. Plots that follow will characterize the nature of the "ramp" from $R1$ to $R2$. Collective data shows that after ≈ 4 yr of spin time, both inner and outer pack plateaus have a mean "ESCA" thickness ≈ 2.6 nm (standard deviation = 0.5). The overall inner regression of plateau decline is $h(nm) \approx 3.2\text{-}0.13\ T(\text{yr})$ with outer-pack surfaces approaching this line a few angstroms faster than inner pack surfaces. The thickness of the lubricant molecule is ≈ 0.7–0.8 nm. In comparison, the overall average height of "ESCA" measured surface lube thickness at ID "depleted" annuli is only 0.42 nm.

Fig. 3 shows the radial location of depletion, $R1$, versus HDA spin time in months. Depletion to a surface lube level near 0.5 nm thickness becomes noticeable after about one year of spin on this type of particulate media and then subsequently widens radially at an average rate near 6.3 mm per year. After 5 years of spin, the depletion annulus

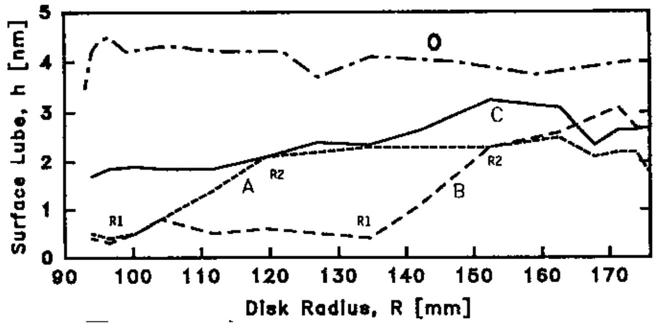


Fig. 1. Surface lubricant thickness "ESCA(R)" versus disk radius after 77 months spin time. 0—Initial lubicant levels. A, B—final outer pack surface profiles. C—Inner pack profile.

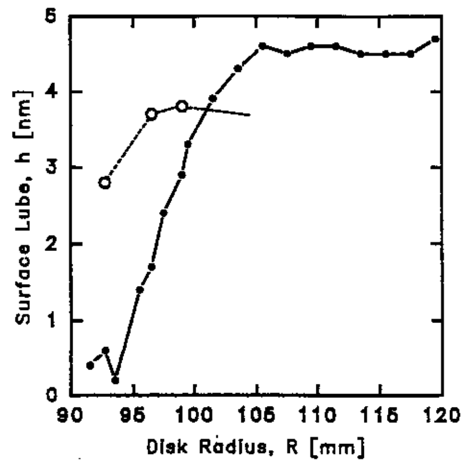


Fig. 2. ESCA surface lube thickness versus radius "ramp" detail at 14.5 months spin time on outer pack surface (o—inner pack).

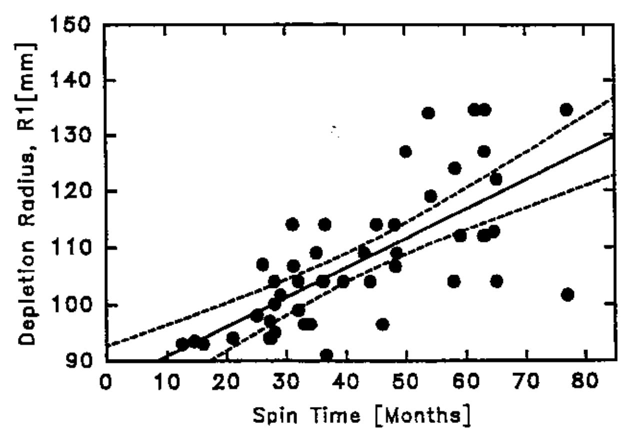


Fig. 3. Inner annulus lubricant depeletion radius versus spin time, $R1(T)$.

has an average radial width of nearly 25 mm. The short-dashed curves show a 95% confidence level surrounding the linear regression line of the data. Lube depletion does clearly grow with spin time. However, the regression coefficient is only $r = 0.7$ and implies that spin time, T, is not the only parameter determining the degree of disk lube migration. Fig. 4 shows how well the final ramp radius, $R2$, correlates with the initial ramp radius, $R1$ ($r = 0.955$). $N = 49$ outer pack disk surfaces were analyzed by point-line ESCA to estimate the radii $R1$ and $R2$ for each surface. In most cases, the ramps had clear resolu-

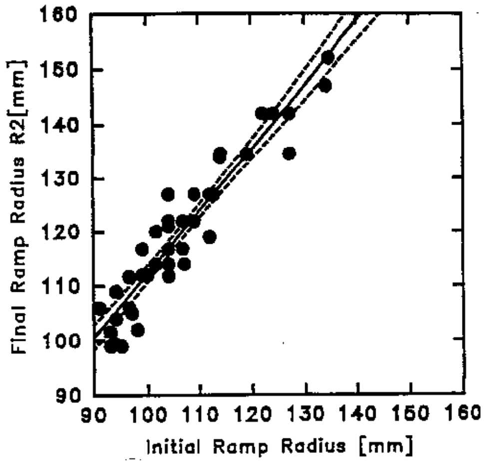


Fig. 4. Correlation of outer ramp radius to inner ramp radius, $R2(R1)$.

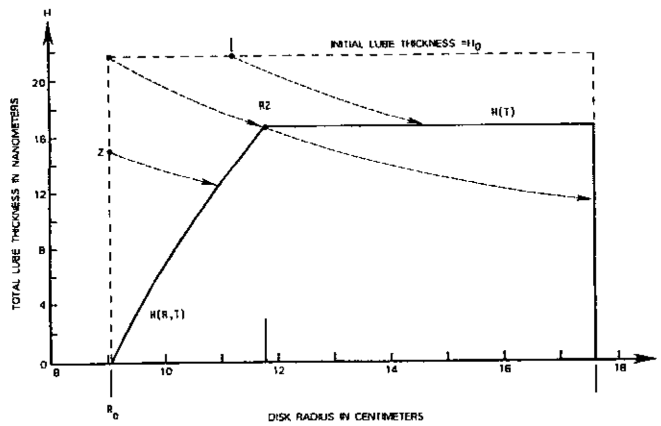


Fig. 5. Total lube thickness profile, $H(R, T)$, for wind driven migration with zero slip length.

tion. The slope of the regression line is higher than one, and the ramp radial extent from an initial radius of $R1$ increases from an average of 10.7 mm at 90-mm disk radius to 17.0 mm at a radius of 130 mm. The ID contact angle after one yr spin time is $\approx 2 \times 10^{-7}$ or ≈ 0.04 seconds of angle. The longest spin examined in Figs. 3 and 4 is 77 mo. In classical models of viscous lube flow away from the boundary ID initial condition (no lube interior to $R = R_0 \approx 90$ mm), both internal and air stress driven migration modeling shows that the point $(R, h) = (R_0, 0)$ would be a "pivot point" of the development profile $h(R, T)$. An example showing computed total lube thickness, $H(R, T)$, for "wind driven migration" is shown in Fig. 5. The fact that the experimentally observed radius $R1$ increases with time implies that $(R, h) = (R_0, 0)$ is not a pivot point. $R1(T)$ develops "as if" liquid interface slippage were occurring.

The interior pack (magnetic read/write utilized) disk surfaces lose ESCA thickness quickly during the first year of spin from an initial $h \approx 4$ nm of surface lube down to an average of ≈ 3.1 nm, but they then only lose ≈ 0.16 nm for each additional year of field spin. Thicknesses between 2.2–2.8 nm appear to be a crudely stable level for many years. The surface lube profile is generally flat; but, at about 5 years of spin, there is a poorly developed "ID cut" (a weak ramp from ~ 1.51 nm at $R = 94$ mm to 2.34 nm at $R = 99$ mm). Air flow measurements performed on our HDA's (Fig. 6) show that the inner pack ID air moves as a solid corotating body for at least several centimeters outwards from the spacer rings. Therefore, little differential air movement exists to produce air shear stresses for this case. At larger radii, differential air flow becomes more significant. In contrast, for the outer pack surface, ID tangential air flow is roughly half of the disk ID rotation speed. We therefore expect significant ID lube migration on outer pack surfaces and almost no cutaway of lube from the ID of inner pack surfaces. Observation shows that this is indeed the case.

B. Bulk Lubricant by Solvent Rinse

Another measurement method for monitoring the state of disk lubrication is the rinsing of disks with freon or

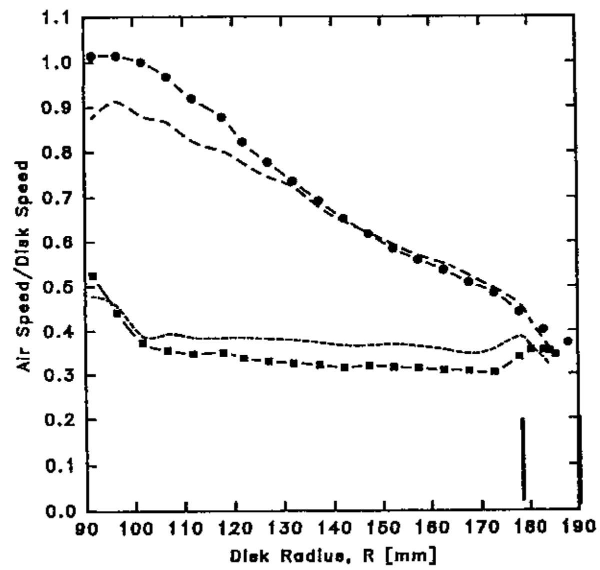


Fig. 6. Measured tangential air velocity/disk rotational speed versus disk radius. Dashed curves are "arms at ID."

with Fluorinert(3M) solvent and then weighing the removed residue which remains after boiling off the solvent. This may be performed per disk surface or by sequential annuli from OD to ID. If adequate coating exists on the OD and OD-chamfer of the disk, then lubricant will migrate to there; and the accumulation can be measured by a "rim strip." Rim lubricant is measured by rotating the outer 7 mm of rim of the disk through a watch glass containing freon. Fig. 7 shows the growth of rim lubricant for $n = 137$ such measurements on C2 coatings. The rim weight is divided by 2 to represent each disk surface. Rim weight increases during the first 30 mo of HDA spin and then stabilizes at ~ 1.18 mg/surface (standard deviation $= 0.24$). At the resolution of this test, differences between interior pack and outer pack disks cannot be seen. The remainder of the disk may be "annular stripped" by freon rinsing of three annuli of equal area. All of these annuli lost weight over the first 2 yr of spin. This technique is not noted for its accuracy; but crude averages after 3 yr of spin show that the inner annuli have lost an average ≈ 0.7 mg of lubricant. Since the outer rim has gained roughly 0.8 mg of lubricant, migration to the rim

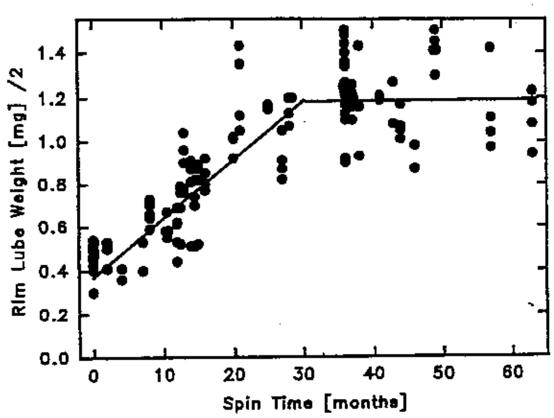


Fig. 7. Accumulation of lubricant on outer disk rim due to migration versus HDA spin time.

has indeed occurred; and evaporation of lubricant must not be very significant. This is interesting because modeling indicates that evaporation should be very significant within the first month of spin [9]. Our independent theoretical calculations also predict rapid evaporation within a month; so we concluded that at least one of our assumptions must be in error. We believe that thin film surface lubricant effective vapor pressure must be well below bulk level vapor pressure. It has been shown that very thin lubricant films have molecules which are not in a bulk state but which instead lie flat on the surface due to strong attractive forces [12].

C. Total Lubricant Thickness by FTIR

A more precise measure of total lubricant thickness, H, integrated throughout the binder is "FTIR" (Fourier transform infrared) analysis of perfluorinated polyether lubricant peaks [13], [14]. In FTIR, infrared light which is polarized parallel to the plane of incidence and which has a high angle of incidence ($\approx 70°$) is used for maximum absorbence in the thin disk coating [14]. Krytox or Y-lubes have strong absorbences between 1100–1350 wavenumbers, but we use a smaller side band near 985 cm^{-1} for analysis. Representative stripped disks are used for background subtraction from the peak near 985 cm^{-1} on lubed disks being evaluated. Background subtraction near these wavenumbers is more reliable than at the larger absorbence peaks. Calibration of FTIR lubricant at the used absorption peak has to be performed statistically against freon surface strips or annular strips. Applicationof this technique also shows loss of lubricant on the disk surface over time and has enough resolution to distinguish between interior pack and exterior pack surfaces. In the inner depleted regions of outer pack disks, FTIR shows about 6 ng/mm^2 average remaining lubrication. This corresponds to an integrated thickness of 3.2 nm of lube throughout the porous coating down from an original 24 nm total thickness.

D. ESCA Versus FTIR Lubricant Thicknesses

Lubrication is often applied by one or more sequences of an atomized spraying during disk rotation which is followed by buffing operations with an absorbent cloth. There is a maximum weight of lubricant which can be absorbed in the binder before the surface lubricant thickness becomes too high and hence too easy to buff off. Let this "saturation lube" have a thickness, H_{sat} [nm].

One method of characterizing particulate media is to generate "lube filling curves" (or "saturation curves") by plotting ESCA versus FTIR for $h(H)$. These curves usually reveal a middle linear growth region which then breaks into a steep growth region until saturation occurs. We often approximate these curves by a tangent function but do not have the resolution to see its curvature clearly. The underlying physics which produces these curves is currently unclear. Since the curve can depend on the method or number of sequential lube applications, $h(H)$ is not a unique equilibrium and must vary with some additional parameter. We might then expect the long spin migration $h(H)$ profile to be different. In practice, the high end of the initial tangent type curve is sometimes avoided because it can lead to excessive lube stiction. The middle or "working region" of the saturation curve is well approximated by a straight line of the form,

$$h = h_0 + mH/H_{sat}. \tag{1}$$

Our "C2" type media would use "ESCA" $h_0 \approx 3$ nm, $H_{sat} \approx 80$ nm, and $m \approx 3.7$ over a domain $0.08 < (H/H_{sat}) < 0.86$.

One of the virtues of porous media is the ability of lubricant to replenish the surface after a migration loss or wearing of the surface occurs. Replenishment is easily demonstrated by argon etching of surface lubricant and observation of ESCA increase from a few minutes to an hour after the etching (some bias exists, however, because of the unusual activation of the surface produced by this test). Let $P = dh/dH =$ the slope of the long spin migration curve $h(H)$ at some disk radius. If a thickness, dh, is removed from a thin annulus of the disk surface; replenishment will result in a net loss in the value of $dH = dh$ such that the net surface loss is eventually $P \times dh$ instead of dh. The effect of the factor P is then to retard surface lube thickness loss. In this sense, one could model only the lube thickness above the disk surface and treat the lube viscosity or the migration driving force as if it were varying with the thickness of the lube in such a way as to duplicate the effect of the replenishment factor, P.

The "depletion curve" lies below the initial lube filling curve. Fig. 8 shows a collection of $n = 77$ pairs of "ESCA versus FTIR" data for sample disks which have sustained years of migration. The sample radii used are $R = 96$ mm, 109 mm, and 137 mm; they are merged because little noticeable difference is seen between them. The dashed line represents the approximate initial filling curve, and the initial lubricant state is at $h \approx 4.1$ nm and $H \approx 24$ nm on that curve. No initial data exists for $H <$ 9 nm, but the filling curve has to drop to the origin because the total lube thickness includes the surface lube thickness. Almost all of the data points lie below the initial dashed line. A possible reason for this is that migration replenishment cannot occur until a thickness loss gra-

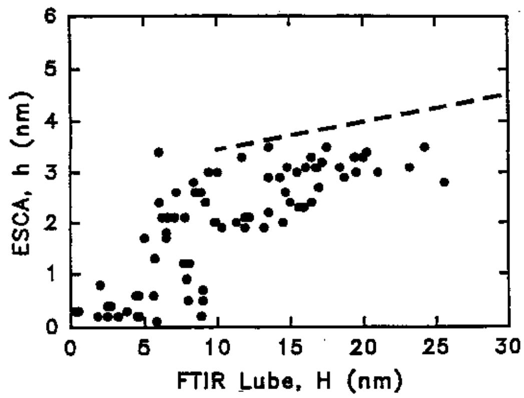


Fig. 8. ESCA versus FTIR lubricant depletion scatterplot.

dient has been established from the top down. In contrast, the initial lube filling was from the supply at top to the initially empty porosity. This sort of decline of $h(H)$ profile has been noted on some other types of media as well as the C2 type reported here. The exact shape of the depletion curve is not clear from the scatter plot data supplied; but a crude approximation, $h = PH$, is possible. In this case, the replenishment factor, P, would also be a constant of proportionality.

III. Mathematical Model

Our initial lubrication replenishment assumption will simply be that the relationship between surface versus total lubrication thickness during migration is given by: $h = PH$. The fluorocarbon lubricant is characterized by a density "ρ," a viscosity "μ," and a kinematic viscosity, "ν" $= \mu/\rho$. The air flowing over a spinning disk will have parameters with subscript "a": ρ_a, μ_a, ν_a. The boundary layer thickness of the air above a spinning disk is a few multiples of $\delta_a = (\nu_a/\omega)^{0.5} \approx 0.2$ mm, where ω is the rate of spin of the disk in radians per second.

The air (or "wind") shear stress is given by [1], [4], [15].

$$\tau_w = \tau_{rz} = R(\mu_a \rho_a \omega^3)^{0.5}/2 = \delta R \omega^2 \rho, \tag{2}$$

where μ_a = viscosity of air, R = disk radius, and $\delta = (\nu_a/\omega)^{0.5}\rho_a/2\rho$. As recently as 1983, the majority of DASD corporations believed that the primary driving force for disk surface lubricant migration was inertial. The variable δ was then selected for ease of comparison to inertially driven migration. A physical interpretation of the term δ is that it would be the thickness of a layer of fictitious lubrication above the actual lubricant surface which would duplicate the effect of wind shear stress in the real lubricant-to-air surface via inertial forces only. A typical value of δ is 60 nm, and a typical value for the radial (and also the circumferential) wind shear stess is $\tau_w \approx 2.3$ Pa (for temperature $\approx 44°$C, $R \approx 110$ mm). Couple this to the shear stress of the surface lubricant by the boundary condition, $\tau(z = h) = \tau_w$ (where altitude $z = 0$ refers to the binder/lube interface). Then, following a procedure similar to that outlined in the well known rotational flow model of Emslie [18], we obtain a partial differential equation for both inertial and wind shear stress driven lubricant migration,

$$\frac{\partial h}{\partial T} + kR\frac{\partial h}{\partial R}(h^2 + h\delta) = -k\left(\frac{2h^3}{3} + h^2\delta\right), \tag{3}$$

where $k = P\omega^2/\nu$ now includes replenishment, and "real P" $\approx P/f$ (the ESCA correction factor). Introducing replenishment effects in this way (replacing a dh loss by Pdh for constant slope P) is verified by an independent finite difference computation which yields the same final lubrication profiles as those calculated from the equation (3). This equation can be solved in closed form from $T(h)$ and $R(h)$ and yields lube thickness profiles for $h(R, T)$ in which the inner boundary wall becomes a ramp which pivots about the fixed point $(R, h) = (R_0, 0)$. This fails to conform to experimental observation, however. The experimental migration profiles would suggest that the conventional "no-slip boundary condition" might need to be relaxed. We then assume that there is an effective velocity of lubricant slip at the lube/binder interface which is proportional to the local lubricant velocity gradient,

$$v(z = 0) = B(dv/dz). \tag{4}$$

The proportionality factor "B" is called the slip length or extrapolation length. Volume flow rates, $Q = d\,\mathrm{Vol}/dT$ (per unit of circumference), may be calculated by successive integrations of Navier Stokes equation to yield:

$$Q = \frac{1}{\mu}\left(\rho R\omega^2 - \frac{\partial p}{\partial R}\right)\left(\frac{h^3}{3} + h^2 B\right) + \frac{\delta R\omega^2}{\nu}(h^2 + hB). \tag{5}$$

The last term represents wind driven flow with interface slip, and p is the pressure which is the disjoining pressure ($\Pi \approx -A/(6\pi h^3)$) plus the meniscus pressure ($-\gamma/a(R)$). $a(R)$ is local radius of curvature, γ is the surface tension, and A is the Hamaker constant (typically $\approx 10^{-19}\,J$ [12]). The validity of using "bulk level" surface tension values, and the estimate for the Hamaker constant and for the power of h valid for our ultra thin film situation is currently unclear. Because of this and also the complexity that would be introduced, the pressure term is deleted from the equations which follow. The flow rate Q is used to produce partial differential equations from the equation of continuity

$$\frac{\partial h}{\partial T} = -\frac{1}{R}\frac{\partial}{\partial R}(RQ). \tag{6}$$

Finite difference equations derived from (6) were programmed for wind driven lube migration and also for disjoining pressure gradient flow. The numerical modeling was used to check analytical models but also has the advantage of being more flexible (e.g., being able to use arbitrary initial lube profiles). But, analytical solutions also have their worth. It is interesting to examine the relative strengths of inertially driven migration to that of wind driven migration.

$$Q(\mathrm{wind})/Q(\mathrm{inertia}) = \delta(B + h/2)/(hB + h^2/3). \tag{7}$$

For surface thickness ESCA near 4 nm and $0 \leq B < 3$ nm, this ratio is 23–30. Thus, wind driven migration strongly overwhelms inertially driven migration for the case of thin liquid films. We also note that a flow rate near the observed 0.3 mg/y is close to the flow rate predicted by the simplest wind flow model slip ($Q \approx 0.5$ mg/y for an average ESCA near 3.6 nm). The agreement between theoretical and measured flow was even better for the next generation of media ("C3"). Therefore, wind shear stress alone does have the ability to drive lubrication at observed rates, and the concept of inertially driven migration may be deleted. Since inner pack surfaces have plateau level ESCA nearly as high as plateau level ESCA on outer pack surfaces, it is suggested that inner pack driving forces must not differ drastically from that of the wind shear on outer pack disks for radii away from the ID. Even though the spacing between corotating disks is only 7.6 mm, that spacing divided by δ_a is ≈ 38, which is not a small number. Even though the circumferential air shear stress should be strongly different for outer versus inner pack disks, the radial shear stress might not be strongly different for radii away from disk ID.

Another useful parameter will be the "efficiency, E," expressed as the experimentally deduced shear stress ratioed to the value of wind shear stress of an ideal free disk with no hole. E will include the air shear stress coupling to the lubrication layer and various factors which may impede lube flow across the disk. If the outer pack disk surface were "free," then we would expect the inner annulus of the disk to have laminar air flow, but the outer annulus of these large disks should have turbulent air flow. The transition radius may occur near $R \approx 110$ nm where the Reynolds number $(R^2\omega/\nu_{\text{air}}) \approx 3 \times 10^5$. The turbulent radial shear stress is given by [17]

$$\tau_r[Pa] = 0.00225\rho_a(\alpha R\omega)^{7/4}\nu_a^{1/4}(1 + \alpha^{-2})/\delta_a^{1/4}, \quad (8)$$

where $\alpha = 0.526$, and $\delta_a = \alpha(\nu_a/\omega)^{1/5}R^{3/5}$. In the outer annuli of the disk, this shear stress is expected to be $\sim 50\%$ higher than the laminar radial air shear stress. The multiplicative enhancement above laminar values is nearly linear with disk radius ($\approx 0.267 + 0.00782\,R$[mm]). This means that $E(R)$ may rise gradually with disk radius. Therefore, even on the outer pack surfaces, the "efficiency" could be a variable rather than a constant. For inner pack disks, we also assume that the efficiency of wind driven migration is approximately linear, $E = E(R) = \phi + \psi R$. The radial air shear stress of inner pack corotating surfaces is not yet well characterized in current literature. However, at least one study indicates that the radial wind shear stress at the surfaces of corotating disks might be significant (perhaps 30% of the outer pack "free" surface values [18]). Incorporation of the parameters B and $E(R)$ leads to a partial differential equation of the form:

$$\frac{\partial h}{\partial T} + \frac{\partial h}{\partial R}[\delta EkR(B + h)] = -\delta k[E(2hB + h^2) + R\psi(Bh + h^2/2)]. \quad (9)$$

The initial condition (IC) of this equation is usually (initially flat surface lube) $h(R, 0) = h_0$ for $2R_0 < 2R < OD$ and (ID-IC) $h(R, 0) = 0$ for $R < R_0$. One of the characteristic curves of (9), the development curve from a point (R_0, h_0) to (R, h)), can still be found in closed form,

$$E(R)R^2/(E(R_0)R_o^2) = [(h_o^2 + 2h_0B)/(h^2 + 2hB)]. \quad (10)$$

But the others require numerical methods. For the outer pack surface approximation, E is constant, we can obtain $h(T)$ in closed form:

$$h(T) = \frac{h_0}{(h_0/2B + 1)\exp[2kEB\,\delta(T - T_0) - h_0/2B]}. \quad (11)$$

In the limit as $B \to 0$, this equation reduces to the easier form: $h(T) = (h_0/1 + PEkh_0\,\delta T)$. Then, in the asymptotic limit as $T \to \infty$, we obtain $h \approx 2\mu/[PE\omega^{3/2}T(\rho_a\mu_a)^{1/2}]$. With the exception of the factors P and E and the square root of air density, this equation is the standard form given in [19]. Again, note that this form has limited usefulness because it fails to deal with replenishment and the existence of the ID hole. For outer pack surfaces, the two equations (10), (11) suffice to calculate the approximate ramp and plateau level profiles $h(R, T)$. Fig. 9 shows a sample computed output profile for 40 months spin time with $E = 18\%$, $B = 3$ nm, and $h_0 = 4.1$ nm. These values for E and B provide a best fit to the overall data profile set, and effective efficiency E had a plausible value. In this sense, mathematical modeling is reasonably consistent with real measurement. The ramp-to-plateau "vertex" should lie along a characteristics curve of the partial differential equation. To obtain the low depletion level of final ESCA, the model also made the ad hoc assumption that a residual 0.5 nm (less than monolayer thickness) of the surface lubricant was "bound" and did not participate in migration. If we let $E = E(R)$, then the time from h_0 to h is evaluated by the integral,

$$T = \int_{h_0}^{h} \frac{dh/(\delta k)}{(h^2 + 2hB)[E(R) + R\psi/2]}. \quad (12)$$

The term in brackets is evaluated at each increment of h using (10). Let the original lubricant boundary (the ID IC and initial flat lube surface) be a set of points $\{(R_0, h_0)\}$. After an elapsed spin time, T, (10) and either (11), or (12) yield the final lube boundary, $\{(R, h)\}$. Computation shows that on a uniform disk with initially flat surface lubrication, if $E(R)$ ramps up from ID to outer diameter (OD), then the final "plateau" outer annulus after the ID ramp should tilt down from ID to OD. This downslope is not observed experimentally. Long spin $h(R)$ curves are usually flat with an occasional rise to the OD. It is therefore possible that the shrouds about the spinning disk pack modify our expectations so that E is approximately constant. Another possibility is that plateau levels

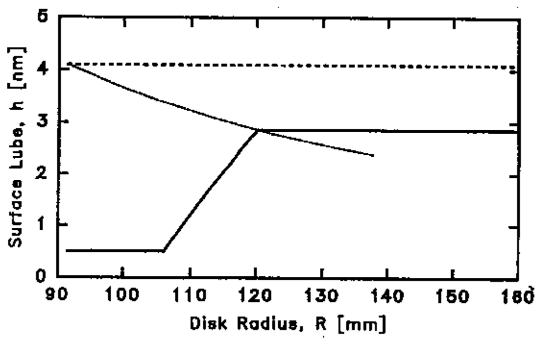


Fig. 9. Lubrication radial profile output from mathematical model.

have gone asymptotic within a few years, so that the migration approach is no longer visible. Another important discrepancy is the ramp-to-plateau "vertex" development with spin time. All wind shear and inertial models predict a gradual motion of the vertex ($R = R2$) across the disk with time (e.g., the dashed development curves of Figs. 5 and 9). Real data shows that the vertex actually falls with more initial steepness from higher to lower ESCA within a shorter radial distance. For other types of disks with higher initial ESCA, this fall has sometimes been seen to be more dramatic and steeper than the data presented here. The combination of this result with the decline of $h(H)$ profiles indicates that the first few year's response of ESCA h values involves a "vertical migration" component in which surface lube appears to fall into the disk porosity. This new effect competes with "in-plane" radial lube migration. It is possible that this effect is a consequence of the vertical gradient switch from the filling versus the depleting of disk lubrication from the surface. These unexpected new observations provide directions for future research.

IV. Conclusion

The fact that lubricant is lost from the data zones of disks and accumulates on the disk rims (and sometimes on baseplates and shells in the planes of the disks) means that radial lubricant migration is actually occurring. The fact that we can account for most of the displaced lubricant means that lube evaportion is not a dominant loss mechanism. The fact that outer pack disks lose more inner annuli lubricant than do corotating inner pack disks means that the driving mechanism near disk ID is stronger on outer pack surfaces. This would be expected because "core" air rotation between interior pack surfaces tends to approximate disk rotation speed more closely. The fact that nearly complete lube depletion occurs on the inner annuli of outer pack disks may imply that conventional fluid dynamics is not adequate. Outer pack lubricant migration evolves as if interfacial slippage were occurring. The primary characteristic of the final migration state is a well developed ramp separating an inner depleted annulus from an outer annulus with a plateau level of lubrication. Disk porosity provides a reservoir for lubricant replenishment to the disk surface which effectively retards the loss of ESCA surface lubricant with time. Model equations were devised which included the effect of proportional replenishment, interface slip, and wind shear stress driven migration. The model using parameters for driving efficiency, E, and slip length, B, was matched to long spin experimental data. Although "E" was intended to be a free parameter in this model, its fit value on outer pack disks was in an acceptable range. This implies that wind driven migration is plausible as the driving factor for outer pack disk lubricant migration. The six-year examination yielded several unexpected results. Additional research should explore the nature of the ESCA drop between the lube filling curve and the depletion curve $h(H)$, the initially precipitous decline of the ramp-to-plateau vertex, and also the possibility of deriving the observed ESCA profiles $h(R, T)$ by a method which avoids interface slippage. If interfacial slip is indeed occurring, its underlying physics needs to be elucidated. Some steps in this direction have already been taken [8]. The mean radial air shear stress values, $\tau_w(R)$ at the surface of inner and outer pack surfaces also needs to be better characterized. Ultimately, more of the ignored factors mentioned in the introduction should be included in modeling in hopes of a more comprehensive explanation of the simple ramp profile evolution on outer pack disk surfaces.

Appendix I

Commercial epoxy-phenolic gamma-ferric-oxide particulate media may be further classified by the type of plastic additive incorporated into the binder. Roughly 10% of the initial organic binder weight may be either the chemical agent PVME (polyvinyl methyl ether [20], [21]), Butvar (polyvinyl butyral), or the thermoplastic CAB (cellulose acetate butyrate [22]). The magnetic coatings with these ingredients may be called type P, type B, or type C. Coatings made with PVME tend to have a high "porosity"—a high capacity to absorb fluorocarbon lubricants. Type B coatings tend to have very little porosity because Butvar is an anti-foaming agent for disk binders. Type C coatings tend to be intermediate; but, under the right conditions they may have as high a porosity as type P coatings. In this paper we examined type C coatings on 3380-type "large disks" (356-mm [14″] diameter). C type disks have evolved over many years from a subcategory which could be labeled as "C1" to "C2" and then "C3." The very long spin time study evaluated here is primarily devoted to type C2—the C type disk at its middle stage of development.

The 356-mm outer diameter size disk has an inner diameter of 168 mm and brown particulate coating applied from a radius $R \geq 91$ mm. The media is spin coated onto an aluminum substrate which is 1.9-mm thick and has no conversion coating. In a head-disk-assembly (HDA), the disk spins at $\sim 3{,}600$ r/min. The outer-pack surfaces of the HDA are separated from the HDA baseplates by a distance of ~ 16 mm. The inner-pack disks are corotating with a spacing of ~ 7.6 mm. The HDA air flow measurements shown in Fig. 6 were performed on an older type HDA ("8650") which also used 356-mm disks, 3,600 r/min spin rate, 7.6-mm disk separation, and similar enclosure spacings. Fig. 6 is a modified form of mean velocity graphs presented in [23]. At a typical operating temperature near 44°C, the type "Y" fluorocarbon lubrication (polyperfluoropropylene oxide) has a kinematic viscosity near 400 mm^2/s. High viscosity lubricants have slow migration rates. Nonbonded Fomblin type "Z" lubricants (a copolymer of tetrafluoroethylene oxide with difluoromethylene oxide) were previously rejected by us largely because their viscosities were too low and the migration rates were too fast. The binder thickness of the C2 disk has a wedge from ID to OD data zone from ~0.5–1.0 μm. The media "packing fraction (*PF*) is the volume of iron oxide relative to total coating volume and is typically 29%. The porosity of the coating is defined by the "saturation capacity" of the applied lubrication ratioed to the total volume of the coating and has a value near 12% (later type C3 and type P media achieved porosities near 18%). To a large degree, porosity is determined by the PF of the iron oxide, its orientation, chemical additives, and the durations and temperatures of oven baking of the coated media. We have not yet found a method for reliably characterizing sizes of pores on actual disk coatings. Mercury porosimetry on large samples of baked media (grams rather than milligrams) indicates pore "diameters" peaked near 0.1 μm, and high magnification scanning electron microscopic analysis also shows some spaces between oxide grains near that size. Porosity is desirable because it allows for a greater quantity of lube on the disk, provides replenishment of surface lubricant to assist protection against surface wear, and retards the decline of surface lube thickness with HDA spin time.

The realization that outer pack disk surfaces were subject to inner annulus lubricant depletion led us to abandon the use of those surfaces for new particulate media data storage products as of 1983. None of the "C2" or "C3" outer pack surfaces discussed here have read/write arms. All of the interior pack surfaces do.

References

[1] S. Middleman, "The effect of induced air-flow on the spin coating of viscous liquids," *J. Appl. Phys.*, vol. 62, no. 6, pp. 2530–2532, Sept. 1987.

[2] W. H. McConnell, "On the rate of thinning of thin liquid films on a rotating disk," *J. Appl. Phys.*, vol. 64, no. 4, pp. 2232–2233, Aug. 1988.

[3] F. B. Derjaguin, G. Strakhovsky, and D. Malysheva, "Measurement of the viscosity of wall-adjacent boundary layers of liquids by the blow-off method," *Acta Physicochimica U.R.S.S.*, vol. 19, no. 6, pp. 541–552, 1944.

[4] D. Peterson, "Wind shear and lube migration," internal StorageTek disk division report, July 1983, unpublished.

[5] S. Kim, J. S. Kim, and F. Ma, "On the flow of a thin liquid over a rotating disk," *J. Appl. Phys.*, vol. 69, no. 4, pp. 2593–2601, Feb. 1991.

[6] G. Teletzke, H. T. Davis, and L. E. Scriven, "How liquids spread on solids," *Chem. Eng. Comm.*, vol. 55, pp. 41–81, 1987.

[7] J. H. Hwang and F. Ma, "On the flow of a thin liquid over a rough rotating disk," *J. Appl. Phys.*, vol. 66, no. 1, pp. 388–394, July 1989.

[8] M. Yanagisawa, "Slip effect for thin liquid film on a rotating disk," *J. Appl. Phys.*, vol. 61, no. 3, pp. 1034–1037, Feb. 1987.

[9] H. Salehizadeh, N. Saka, and E. Rabinowicz, "Lubricant evaporation from rigid magnetic recording disks," in *Tribology and Mechanics of Magnetic Storage Systems*, vol. IV (STLE-SP-22), pp. 87–92, 1987.

[10] R. E. Linder and P. B. Mee, "ESCA determination of fluorocarbon lubricant film thickness on magnetic disk media," *IEEE Trans. Magn.*, vol. MAG-18, no. 6, pp. 1073–1076, Nov. 1982.

[11] K. Nishimori and K. Tanake, "Evaluation of the electron attenuation length for the determination of the lubricant overlayer thickness on magnetic disks with x-ray photoelectron spectroscopy," *J. Vac. Sci. Technol. A*, vol. 8, no. 4, pp. 3300–3303, July/Aug. 1990.

[12] C. M. Mate and V. J. Novotny, "Molecular conformation and disjoining pressure of polymeric liquid films," *J. Chem. Phys.*, vol. 94, no. 12, pp. 8420–8427, 1991.

[13] V. Au-Yeung, "FTIR determination of fluorocarbon lubricant film thickness on magnetic disk media," *IEEE Trans. Magn.*, vol. MAG-19, pp. 1662–1664, Sept. 1983.

[14] F. Walder, D. W. Vidrine, and G. Hansen, "The anaysis of lubricants on magnetic disks by Fourier transform infrared spectroscopy," *Appl. Spectroscopy*, vol. 38, no. 6, 1984.

[15] H. Schlichting, *Boundary Layer Theory*, 4th Edition. New York: McGraw-Hill, 1960, pp. 83–87.

[16] A. G. Emslie, F. T. Bonner, and L. G. Peck, "Flow of a viscous liquid on a rotating disk," *J. Appl. Phys.*, vol. 29, pp. 858–862, May 1958.

[17] D. R. Davies, "On the calculation of eddy viscosity and heat transfer in a turbulent boundary layer near a rapidly rotating disk," *Quart. J. Mech. Appl. Math.*, vol. XII, pt. 2, pp. 211–221, 1959.

[18] S. D. Abrahamson, "An experimental investigation of the flow between shrouded corotating disks," Ph.D. dissertation, Stanford University, Chapter 5, April 1988.

[19] B. Bhushan, *Tribology and Mechanics of Magnetic Storage Devices*. New York: Springer-Verlag, p. 576, 1990.

[20] D. Johnson, R. Flores, and M. Vogel, U.S. Patent No. 3,058,844, 16 Oct. 1962.

[21] V. J. Novotny and T. E. Karis, "Sensitive tribological studies on magnetic recording systems," *Adv. Info. Storage Syst.*, vol. 2, p. 137, 1991.

[22] V. A. Lohoff, U.S. Patent No. 4,122,234, 24 Oct. 1978.

[23] V. A. Sandborn, "Evaluation of the mean and turbulent velocities in a storage unit," Colorado State University Aerodynamics Research Memorandum no. 41, Feb. 1979.

David L. Peterson (M'82) was born in Minneapolis, Minnesota, in 1942. He received the B.S. degree in engineering physics in 1965 and the M.S. degree in physics in 1968 from the University of Colorado, Boulder.

Additional study and teaching in physics and mathematics continued at Colorado University during the 1970's. He joined Storage Technology Corporation in Louisville, Colorado, in 1978 and has participated in disk analysis, magnetics, standards, and tribology. He is presently a senior engineer at StorageTek.

Letter

Long Spin Disk Lube Migration—II

David L. Peterson, *Member, IEEE*

Abstract—**Significant improvements in the persistence of disk lubrication thickness profiles are reported from observation over five years of head/disk assembly (HDA) field spin time for a final version of particulate media. This direct long-term monitoring has more validity than previous attempts at accelerated lube performance testing.**

Radial migration of disk surface lubrication was a significant problem for hard disk drive manufacturers in the early 1980's. The impact was especially severe for those producers who elected to utilize outer pack disk surfaces ("free" or noncorotating surfaces which faced out towards the drive casing) for data read/writes. Wind shear stresses would produce a depletion of lubrication at disk inner diameters (ID's) sometimes within a month to a year of field spin time. If heads were used in that area, they tended to "crash." This letter shows that particulate media ultimately achieved a level of highly reliable tribological performance. The goal for our final particulate media product was to improve lubrication surface thickness persistence by increasing the porosity of the disk media and applying more lubrication. These efforts led to an improvement by nearly a factor of four in the reduced rate of surface lubrication loss. The selection of lube levels is always a tradeoff between tribology and stiction. For perspective, we provide a sample long-term ($\sim$one year) profile of lube stiction force development. It is not clear that lubricant migration is a nonissue with current thin-film media. The reliability of the final generations of particulate media was due in part to significant replenishment from subsurface porosity. Although thin-film media use partially bonded lubricants, the media are largely lacking in replenishability. In addition, thin-film drives also utilize the outer pack disk surfaces and hence are highly subject to wind shear stresses.

In a previous paper [1], we studied disk lubrication persistence for an early version of "3380 type" particulate media over six years of field spin time (and finally over eight years—the longest sample was at 105 months of spin time). No such real study of lube thickness profiles had previously been reported for any type of disk media. Studies had been reported for tests using thick surface lubricants, very short spin times, disks without inner diameter holes, or accelerated spin-off testing—but those tests were largely irrelevant to industry needs or flawed because the causes of radial lubricant migration were poorly understood. References to research performed in the area of disk lubricant migration are provided in [1], [2]. Reference [1] demonstrated clearly that "thin" (<6 nm) surface lubricants flow primarily in response to "wind shear stresses" from HDA internal air flows rather than from inertia ("centrifugal forces"). Accelerated testing (which we also performed) was often based on increasing temperature and rotation rate to boost centrifugal forces on surface lubricants and lower their viscosity. For our 356 mm diameter disks, this also had the undesirable effect of altering the laminar versus turbulent air flow profiles. In addition, thin surface lubricants do not have the same properties as bulk lubricants. The result was unreliable scaling of experimental results. However, the obvious flaw of long-term field testing is the long time required. It might take a decade or more to obtain clear and comprehensive results for well-performing media, and several generations of product occur before the test is complete. Such lengthy testing and postmortem analysis were still desirable and should still have been encouraged for improved industry perspective. They should also be encouraged for current thin-film media.

In our previous study [1], we characterized surface lubricant thickness radial profiles by well-known "ESCA" techniques and total lubricant profiles by "FTIR" (Fourier transform infrared), and "annular freon strip-and-weighs." We report the industry conventional "Tau method" for ESCA lube thickness measurements [3] (although we know that "reality" is different [1] and dependent on the type of media). FTIR used a lube absorbence sideband near 985 cm^{-1} [4] and a spot size of several millimeters. The disk pack air flow was characterized at an altitude of 3 mm above the disk surface. The drives had a spin speed of 3620 r/min, computer shop-controlled humidity, and a typical operating temperature near 44°C. Measurement and characterization details are discussed in the comprehensive report [1] along with a mathematical model of disk lubricant radial migration with spin time. An Appendix also discussed the details of the branched and viscous "Y-type" lubricant and the media (epoxy-phenolic gamma-ferric-oxide with cellulose acetate butyrate thermoplastic add—called "C2" for "CAB" type media at its intermediate level of development). In this letter, we show that a desired improvement in lube retention of over a factor of four was produced largely from an additional 50% increase in binder porosity on a new (and final) media labeled "C3." The porosity of this media averaged 18% (defined as "spray-and-buff saturation capacity" of applied lubrication ratioed to total media coating volume). Our products used a "Fomblin Y" (or, equivalently, a Krytox 143-AD polyperfluoropropylene-oxide) type unbonded disk lubrication with a bulk kinematic viscosity of about 400 mm^2/s at drive temperatures near 44°C. The lube quantity averaged about 10 mg/surface, which was later dropped to 8 mg/surface to avoid stiction; and the long spin testing reported here was for disks with the initial 10 mg of surface lube. Disks were polished to a flat wedge near 0.5 μm thickness. The initial "ESCA surface thickness" was about 6.4 nm. We also used a more traditional "PVME" type epoxy media ("P") in another related product, and compared it against the C3 media. Both media had nearly the same initial (and final) lubrication levels and had initial lube quantity at about 73% of saturation quantity. Their final tribological and lubricant performances were essentially identical. Porosity improvements in our primary C3 media resulted from improved binder chemistry, improved oxide loading, and longer baking time and temperature which paradoxically led to a much longer lived, but also softer media more prone to handling damage during assembly. When combined with improved particle contamination control, improved mechanical design, and improved control of magnetic contamination [5], we obtained a 160 year MTBF HDA (for a dual-

Manuscript received December 12, 1995; revised March 25, 1996.

The author is with the Systems Analysis/Tools Departments, Storage Technology Corporation, Louisville, CO 80028-4205 USA.

Publisher Item Identifier S 0018-9464(96)04878-9.

actuator HDA). This MTBF result is easily competitive with that of thin-film media drives (but particulate media cannot be used for very low flying heights).

The initial response of "ESCA surface lube thickness h" during the first few months of spin is a decline of nearly 1 nm. After this thickness loss, a lube concentration gradient normal to the surface is established in the media; and replenishment of surface lube occurs from the subsurface binder reservoir. After the first two years of field spin, the "C3" media "ESCA surface lube thickness" for outer pack surfaces dropped from about 6.2 to about 4.2 nm and had a radial migration flow rate near Q = 1.1 mg lube loss per surface per year. The simplest wind shear lube migration model discussed in [1] predicted this value almost exactly. This is surprising and probably fortuitous. Other measured results showed that about a third of the initial bulk (FTIR measured) lubrication and about a third of the overall surface (ESCA) lubrication were lost on this medium after two years of spin time. This was largely due to the initial surface lube thickness being above the "saturation knee" (above which ESCA thickness climbs steeply with applied lube quantity). After this excess lube was gone, we then entered a slow loss regime with much higher plateau levels of lube than we had on the previous "C2" product. Additional data showed that the plateau regression on inner used surfaces was $h(t)$ = 58.7 Å − 0.385t (time in months). If this trend continued, the main body of the disks would not be subjected to lube thinner than 20 Å for over eight years of spin time. But this result includes the initial rapid ESCA decline, and is hence overly pessimistic. The "P" type media began with 5.7 nm surface lube thickness (σ = 0.47, n = 40) and 11 mg lube/surface (but with a thicker and ramped binder—so that lube is below the "saturation knee"). The plateau level lubrication drops as $h(t)$ ~ 57.5 Å − 0.322 t [months].

We desired a metric for comparing the lubrication performance of "C2" versus "C3" media. In [1], we saw that the easiest short-term test for lube longevity was the monitoring of the decline in ESCA surface lube thickness at the inner diameter (ID) of "outer pack" disks. In Fig. 1, we show an ESCA surface lube thickness profile at 55 months of field spin time. For this sample, four outer pack disks showed ESCA surface lube ID thickness of 0.6, 0.7, 0.8, and 1.0 nm (considered to be danger levels). Depletion of lube from the deliberately unused outer pack disk ID's could occur after only one year of spin time for older products [1], but at about 4.5 years for the final "C3" disk formulation. A comparison of mean performances showed that C3 was an average of 4.7 times better than C2. This also implies that we would need to monitor this product for more than 20 years to obtain the many details we reported for the C2 disk. Fig. 1 also shows that the used inner pack corotating surfaces have a higher and more resistant surface lube thickness due to reduced wind shear stresses between corotating disks. Typical lube thickness levels on used surfaces plateau across the disk at about 37 Å surface lube thickness after 55 months of field spin time. The outer surface lube thickness ramp (up to the "plateau" lube level) has a radial width of about 11 mm—the same width previously observed for the C2 media. Fig. 2 shows a set of 43 measurements for "spin time to ID depletion" made at a standard radius of 94.5 mm.

A lubrication-related problem encountered with this C3 medium was our initial desire to fully utilize our increased porosity by applying lubrication "at or above the knee" of the applied lubrication "saturation curve" (of measured ESCA thickness versus milligrams of finally absorbed buffed lubricant). This led to high "head/disk" stiction growth. Fig. 3 shows a plot of striction growth over a year of observation time for those initial HDA's which had high lubrication. Again, no such long-term study had previously been reported in the literature. Stiction forces were measured by torque meter testing on the ground hub nut of the HDA. We applied a slowly increasing torque to the disk pack until it broke free from the head assemblies. These particular HDA's had 75 Å of ESCA

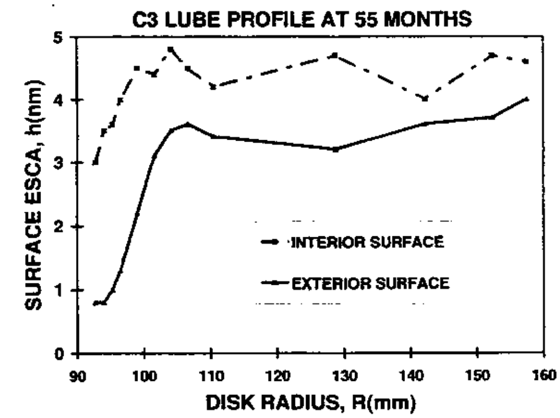


Fig. 1. Surface lubrication thickness profile of sample disk after 55 months of field spin time. Outer peak surface lube has cut away from the disk ID. Corotating surfaces have less migration due to lower radial air-flow shear stresses.

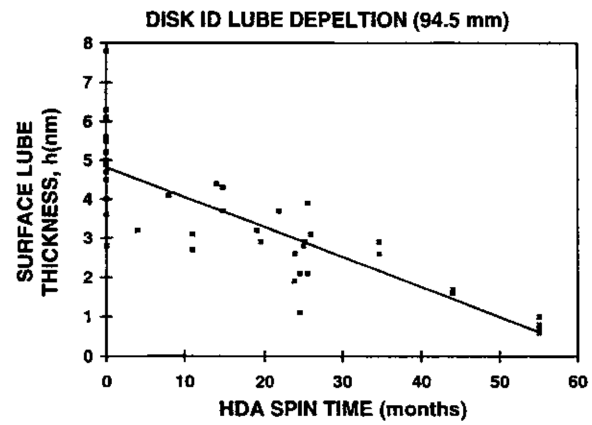


Fig. 2. Loss of inner radius lubrication with spin time. For n = 43 ID ESCA data points, a depletion level of 5 Å will be obtained at 56 months spin time (~95% confidence that the ID depletion spin time is greater than 42 months).

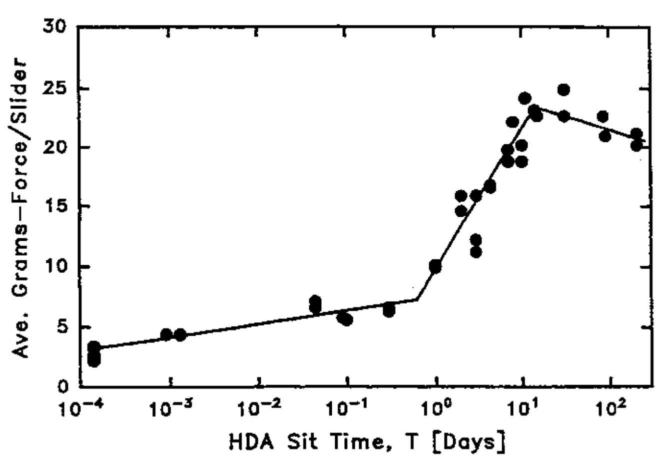


Fig. 3. Tangential stiction forces for two similar HDA's monitored for more than a year. The final data points are for 220 days of room temperature sitting time. Slider load force was ≅ 15 g (0.147 N/slider).

surface lubricant and about 12 mg of applied Y-type unbonded lubricant per disk surface. Stiction was obtained from the calculated force applied to the heads, and was normalized to a "per-slider" basis. We observe three typical regions: low growth, steep logarithmic growth, and finally a "flooding" or "saturation" slow-torque decline near 23 g average force per slider. The solution to this stiction problem was simply to apply less lubrication so that head/disk meniscus growth is discouraged. We were able to select a safe level which still guaranteed adequate lubrication and product longevity (as observed via MTBF and returns for "head/disk interaction").

Acknowledgment

I would like to thank P. Shepard, F. Silliman, G. Burnett, and J. Fechner for their long-term contributions to measurements of lube thickness levels for this medium.

References

[1] D. L. Peterson, "Long spin disk lube migration," *IEEE Trans. Magn.*, vol. 28, pp. 1988–1995, July 1992.

[2] B. Bhushan, *Tribology and Mechanics of Magnetic Storage Devices*. New York: Springer-Verlag, 1990, p. 576.

[3] R. E. Linder and P. B. Mee, "ESCA determination of fluorocarbon lubricant film thickness on magnetic disk media," *IEEE Trans. Magn.*, vol. MAG-18, pp. 1073–1076, Nov. 1982.

[4] F. Walder, D. W. Vidrine, and G. Hansen, "The analysis of lubricants on magnetic disks by Fourier transform infrared spectroscopy, vol. 38, no. 6, 1984.

[5] D. L. Peterson, "Controlling magnetic contamination in the disk drive industry," *Microcontamination* (now *MICRO*), vol. 11, pp. 47–51, June 1993.

Tape Magnetization Waveforms

Dave Peterson
Date: 5/29/01
Preliminary

Several questions were recently posed about what the magnetic field strength would be above the surface of type media and also what the magnetization waveform might look like in the presence of write equalization pulses. These questions are partly addressable by my Fourier Series equations for the "Superposition of arctangent transitions for magnetization." A simple MATLAB model was written ***"magwave.m"*** for plotting output M(x), demag fields Hd(x), and fields above tape Hx(x) and Hy(x). I claim that external fields like Hy are much stronger than previously supposed and that write EQ pulses slightly lower the average peak magnetization for a low frequency waveform (1,7 zeros LF = HF/4) and place ripples at the top of the M(x) profile.

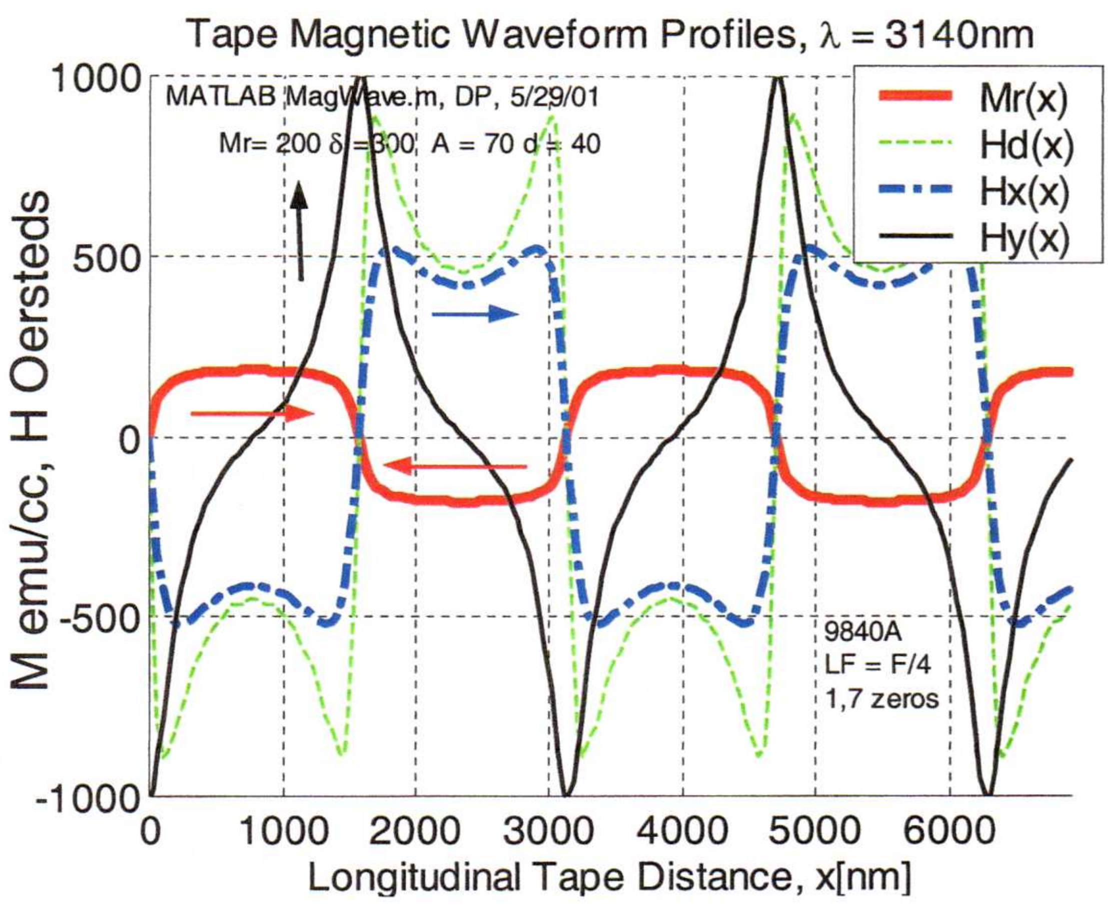


Figure 1. MATLAB *"MAGWAVE.m"* output for a low frequency 840 magnetization pattern. Peak Magnetization is M = 182 emu/cc (for a given Mr = 200 emu/cc and Hc = 1625 Oe. Peak DeMag field is 890 Oe, Hx = 525 Oe, and peak Hy is a huge 1010 Oe (at an altitude of 40 nm above the coating).

SUPERPOSITION OF ARCTANGENT TRANSITIONS FOR MAGNETIZATION.

David Peterson, 5/21/92.

USEFULLNESS OF FOURIER SERIES FOR REPEATED ALTERNATING WAVEFORM.

If a regular series of magnetization reversals is written on a data disk, it is a convention to assume that each transition is Tan^{-1} in shape. The most convenient way to represent the resulting M(x) coating magnetization is by Fourier Series. In opposition, the direct superposition of alternating arcTan (ATN) shapes is more cumbersome, usually converges more slowly, and hence requires more terms. Usually, only a few terms of the Fourier series will suffice for practical applications. The Fourier Transform for isolated ATN's is not defined, so we work with the transform for its derivative, the magnetic charge density, $\rho(x) = -\partial M/\partial x$.

$$M(x) = \frac{2M_r}{\pi}\, Tan^{-1}\left(\frac{x}{a}\right), \quad \rightarrow \quad \rho(x) = -\frac{2M_r a}{\pi(a^2+x^2)}, \tag{1}$$

where "a" is the arctangent transition parameter. The profile of ρ is "Lorentzian", and the "shift theorem" is used to superimpose repeating alternating Lorentzian charges.

$$\rho(x) = -\sum_{1,3,5..}^{\infty} \frac{8M_r}{\lambda}\, e^{-2\pi an/\lambda} \cos\left(\frac{2\pi nx}{\lambda}\right), \tag{2}$$

where λ is the plus-to-plus longitudinal peak separation wavelength. This series is then integrated to get back to the desired M(x) repetition.

$$M(x) = \frac{4M_r}{\pi} \sum_{1,3,5..}^{\infty} \frac{e^{-2\pi an/\lambda}}{n} \sin\left(\frac{2\pi nx}{\lambda}\right). \tag{3}$$

This series could also have been "derived" from an assumption that the Fourier transform of an isolated M(x) transition is $M_r\exp(-2\pi sa)/i\pi s$. The illegitimate divergent contributions would cancel when the alternating series is superimposed. In place of the exponential in the summation, we could use just "1" for the simpler "square wave" pattern, $\mathrm{sinc}(nA/\lambda)$ for ramp (trapezoidal) transitions, or a more complicated $\{(a\pi^2 n/\lambda)/\sinh(a\pi^2 n/\lambda)\}$ for alternating series of hyperbolic tangent transitions.

The use of this series assumes linearity (but we know that the write processes is intrinsically non-linear). For inductive read voltage output, it is the steepest part of the magnetization transitions that counts, and that fact makes this approach generally acceptable. If we wanted to know the actual peak values of recording magnetization, however, then this approach would be more dubious. Real transitions are not perfectly arctangent. For thin film zig-zag transitions, an error function may be more appropriate. The choice of the functional shape of the transitions has a strong affect on peak magnetizations when an alternating series of transitions is superimposed. When bits get crowded close together, the arctan shape would lead to a much stronger attenuation of interbit magnetizations than would the tanh function.

The "Fourier transform generation of a Fourier series" approach is also useful for superpositions of alternating isolated voltage pulse outputs. For example, if the isolated pulse line shape is Lorentzian [$V = V_0\, b^2/(b^2+x^2)$], then the Fourier approach yields the series:

$$V(x) = \frac{4\pi b V_o}{\lambda} \sum_{1,3,5}^{\infty} e^{-2\pi b n/\lambda} \cos\left(\frac{2\pi n x}{\lambda}\right). \qquad (4)$$

The series for "MFM all-ones-- 2F" for pole strength $\rho(x)$ can be used to calculate coating mid-plane demagnetizing fields as a much easier approach than multiple superpositions of the more conventional Potter's expression:

$$H_D(x) = 8M_r\, Tan^{-1}\left(\frac{x\delta/2}{x^2+a^2+a\delta/2}\right), \qquad (5)$$

where δ is the coating thickness. We simply integrate the expression $4\rho \mathrm{ATN}(\delta/[2(x'-x)]$ over all x. (The answer should then properly use x', but we retain x).

$$H_D = -16M_r \sum_{1,3,.}^{\infty} e^{-2\pi a n/\lambda} \sin\left(\frac{2\pi n x}{\lambda}\right)\left(1-\frac{e^{-\pi\delta n/\lambda}}{n}\right). \qquad (6)$$

Finally, consider the strength of the magnetic field above written disk data. This can be calculated from $\rho(x)$ using the older "Wallace" approach with $dH_x = 2\delta\rho(x)(x'-x)/r^2 dx$, where the distance r depends on the altitude z above the midplane of the coating. This gives

$$H_x = \frac{16\delta M_r \pi}{\lambda} \sum_{1,3,.}^{\infty} \sin\left(\frac{2\pi n x}{\lambda}\right) e^{-2\pi n(a+z)/\lambda}. \qquad (7)$$

Magnetoresistive (MR) head reading would see the H_y component (which is the same as the above but with a Cosine term in place of the sin term).

The following collection of papers derives these Fourier series expressions and discusses their use.

GENERATING FOURIER SERIES FROM TRANSFORMS:

Dave Peterson, 5/22/92

The Fourier transform of a function f(x) is defined as

$$\int_{-\infty}^{\infty} f(x)\ e^{-2\pi i x s} dx. \qquad (1)$$

(I like Bracewell's convention of having a factor of "2π" in the exponent and no coefficients of 2π outside the integral). A Fourier transform only exists if the integral of |f(x)| from -∞ to +∞ exists [e.g., the FT of ATN(x) then does not exist, but I show that it is possible to make practical use of it anyway]. If f(x) describes some sort of localized "pulse" shape, we will be concerned with forming Fourier Series from superimposing an infinite series of alternating pulses of shape f(x). This is done using the "shift theorem" and the "addition theorem":

$$\int_{-\infty}^{\infty} f(x-a)\ e^{-i2\pi x s} dx = e^{-i2\pi a s} F(s)\ ;\quad \int_{-\infty}^{\infty} [f(x)+g(x)]\, e^{-i2\pi x s} dx = F(s)+G(s)\ . \qquad (2)$$

We separate an "up" pulse f(x) and a following "down" pulse "-f(x- λ/2)" by half a wavelength, λ. The Fourier transform ("FT") of this dipole will be F(s)[1-exp(-iπsλ)], and we label the dipole itself with the symbol, f^(x). We now form a periodic function, p(x) from the replication of the "up-down" dipole at in interval, λ. This is done by convolution with the replicating symbol, "shah" Ш, or p(x) = Ш(x/λ)*f^(x) = Σf^(x-nλ). Convolution of two functions, say f and g, is defined by ∫ f(u)g(x-u)du. "Shah" can be defined by Ш(x) = Σδ(x-n), or Ш(x/λ) = Σδ(x-nλ). Shah is its own Fourier transform. Now, by the "similarity theorem", if f(x) has the FT F(s), then f(ax) has FT equal to F(s/a)/|a|. Therefore, "FT"Ш(x/λ) = λШ(λs), and Ш(λs) = (1/λ) Σδ(s- n/λ) [e.g., see [1] pg 78]. The function p(x) can then be formed by taking the inverse Fourier transform of the FT of the convolution (which is just a product in frequency space). The inverse FT is defined by: f(x) = ∫ F(s)exp(+2πisx)ds. So:

Define p(x) = [Σδ(x-nλ)]*f^(x) = (1/λ)Ш(x/λ)*f^(x) → Ш(λs)F^(s) (3)

Then, "FT" of p(x) = (1/λ)Σδ(s-n/λ)F(s)(1-$e^{-i\pi s\lambda}$), so let s = s_n = n/λ only.

Note that (1-$e^{-i\pi s\lambda}$) = 1 for odd n but equals zero for even n. Therefore, the series can only have odd values of n (1,3,5...). The sum is from -∞ to +∞, so if F(s) is an even real function of s then the exponential in the inverse FT can be replaced by 2cos(2πnx/λ) with summation only over positive values of n. So, finally, we have:

$$FOURIERSERIES = p(x) = \sum_{1,3,5..}^{\infty} \frac{F(s)}{\lambda} \cos\left(\frac{2\pi n x}{\lambda}\right). \qquad (4).$$

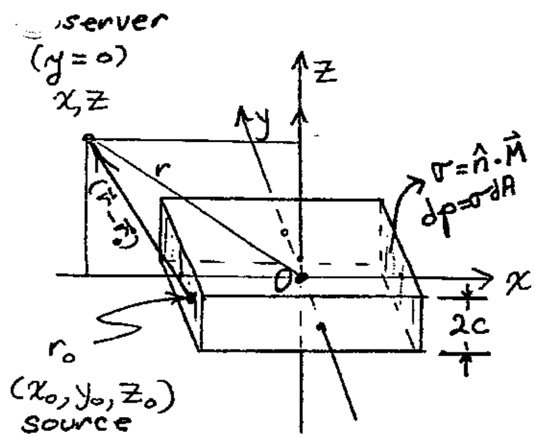
If F(s) were an odd function of s, then the Fourier Series would use $2\mathrm{Sin}(2\pi nx/\lambda)$. The last useful property in Fourier Transforms is the derivative theorem, "If f(x) has the Fourier transform F(s), then f'(x) has the Fourier transform $i2\pi sF(s)$. We are now ready to produce a Fourier Series for any given initial isolated function shape used to form an alternating periodic function waveform. This is useful in read/write in magnetic recording.

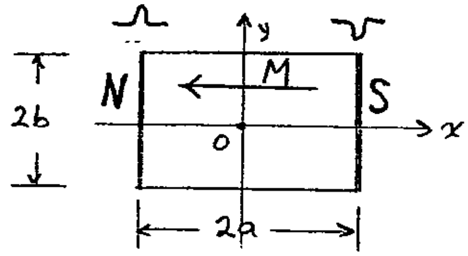
REF:

[1] Ron Bracewell, The Fourier Transform and Its Applications, McGraw-Hill, N.Y., 1965.

EXTERNAL MAGNETIC FIELD FROM A MAGNETIZED RECTANGULAR SOLID

dp 4/9/91





Let a hard magnetic solid have dimensions 2ax2bx2c and magnetization in the -x direction, M(emu/cc).
A surface pole density will then exist at $x = \pm a$, and a magnetic field, H(Oe), will result from the North source at x = -a to the South source at x = +a.

If we let rectangle function $\Pi(x) = 1 \quad |x|<1, =0 \quad |x|>1$.
and $⇅(x)$ = odd impulse pair = $\delta(x+1) - \delta(x-1)$, then $\partial\Pi(x)/\partial x = ⇅(x)$. The pole density function is given by:

$$\rho = M_o ⇅(x/a)\,\Pi(y/b)\,\Pi(z/c) = -\nabla\cdot\vec{M} \quad .$$

We will restrict ourselves to horizontal magnetic fields $dH_x = (x - x_o)dH/|r - r_o|$ where $dH = \rho dV/r^2$ (CGS).

The radius vector from an element of source on a pole face to a location (x,y,z) is $|\vec{r} - \vec{r}_o| = [(x\pm a)^2 + (y-y_o)^2 + (z-z_o)^2]^{1/2}$

A computer could just sum up all the contributions from the area of the pole faces, but it is sometimes convenient to do a little analytical integration before the final computer integration.

$$\int[\delta(x_o+a) - \delta(x_o-a)](x-x_o)dx_o = (x\pm a)\,\mathrm{sgn}(x_o)$$

$$H_x(Oe) = \int_{-b}^{b}\int_{-c}^{c} \frac{M_o\, dy_o dz_o\, (x\pm a)\,\mathrm{sgn}(\pm a)}{(r-r_o)^3}$$

A "Y" integration will be performed first. For maximum resulting Hx field, we will also restrict ourselves to output along a centerline with y = 0.

$$\text{Let } B^2 = (x\pm a)^2 + (z-z_o)^2; \quad (r-r_o)^3 = [B^2 + (0-y_o)^2]^{3/2}$$

$$\text{Dwt } 200.03: \int \frac{dy_o}{r^3} = \frac{1}{B^2}\frac{y_o}{r}\Big|_{-b}^{b} = \frac{2b}{B^2\sqrt{B^2+b^2}}, \text{ so}$$

$$H_x(x,z) = \int_{-c}^{c} \frac{M_o\, dz_o\,(x\pm a)\,\mathrm{sgn}(\pm a)\, 2b}{[(x\pm a)^2 + (z-z_o)^2]\sqrt{(x\pm a)^2 + (z-z_o)^2 + b^2}}, \left(\frac{d(z-z_o)}{dz_o} = -1\right)$$

Actually, even this integral isn't bad. Its closed form expression can be found in Gröbner and Hofreiter's "Unbestimmte Integrale" (springer-verlag 1965- 19a:)

$$\int \frac{dx}{(x^2+c^2)\sqrt{x^2+a^2}} = \frac{1}{c\sqrt{a^2-c^2}} \arctan\left(\frac{\sqrt{a^2-c^2}\,x}{c\sqrt{x^2+a^2}}\right) + C, \text{ for } a^2 > c^2$$

$$H_x(x,z) = -M_o(x\pm a)\,\mathrm{sgn}(\pm a)\left[\frac{2b}{(x\pm a)b}\,\mathrm{Tan}^{-1}\left(\frac{b(z-z_o)}{(x\pm a)\sqrt{(z-z_o)^2 + (x\pm a)^2 + b^2}}\right)\right]\Bigg|_{-c}^{c}$$

$$H_x = -2M_o\left[\mathrm{Tan}^{-1}\left(\frac{(z-c)b}{(x+a)\sqrt{(z-c)^2+(x+a)^2+b^2}}\right) - \mathrm{Tan}^{-1}\left(\frac{(z+c)b}{(x+a)\sqrt{(z+c)^2+(x+a)^2+b^2}}\right)\right.$$
$$\left. - \mathrm{Tan}^{-1}\left(\frac{(z-c)b}{(x-a)\sqrt{(z-c)^2+(x-a)^2+b^2}}\right) + \mathrm{Tan}^{-1}\left(\frac{(z+c)b}{(x-a)\sqrt{(z+c)^2+(x-a)^2+b^2}}\right)\right.$$

Tape Track Distortion

David Peterson

"Long Version"
(8/30/00, StorageTek Symposium'00)
[Presentation with Linus Wang]

Abstract

High density data recording tape has a viscoelastic substrate that can be deformed by the actions of creep, shrink, and environmental factors. The physical distortions of data and servo tracks become increasingly more important and problematic as we increase the number of recorded linear tracks placed on tape. Tape width narrows at outer cassette spool radius due to applied tension and widens near the pack hub due to the combination of compressive circumferential stress and high radial stress. Tape creep development is guided by these stresses and progresses with exposure time. The creep profiles of center-loaded tapes are often referred to as "lips plots" or "bow-tie." Servo position error measurements (PES) over a triplet of servo readers enabled rapid and highly detailed profiling of tape distortion over many tapes. Supporting studies include physical tape characterizations, mathematical analysis, optical width measurements, pack level stress modeling, and responses to high environmental temperature. Severe distortion of tape can affect "servo dynamic range" and lead to servo errors. Distortion can also limit track misregistration margins and contribute to R/W data errors. The development of future tape products with high track density requires strategies to deal with these exposures.

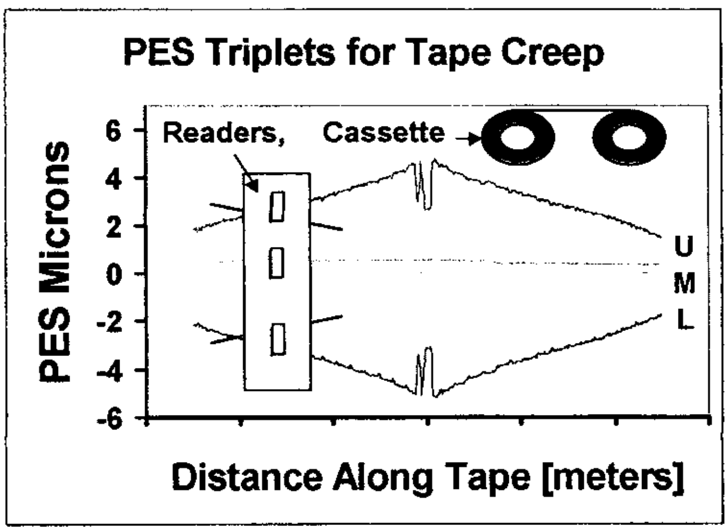


Figure 1: Example of Position Error Signal ("PES") output profiles for a triplet of magneto-resistive servo readers (upper, middle, and lower elements, U,M, and L). The PES Tester traces a full wrap of data from one end of the tape to the other along a two-spool cassette.

Introduction

As a tape is stretched under tension, it elongates elastically in the direction of the stretching force and also narrows its width in the transverse direction. Servo and linear

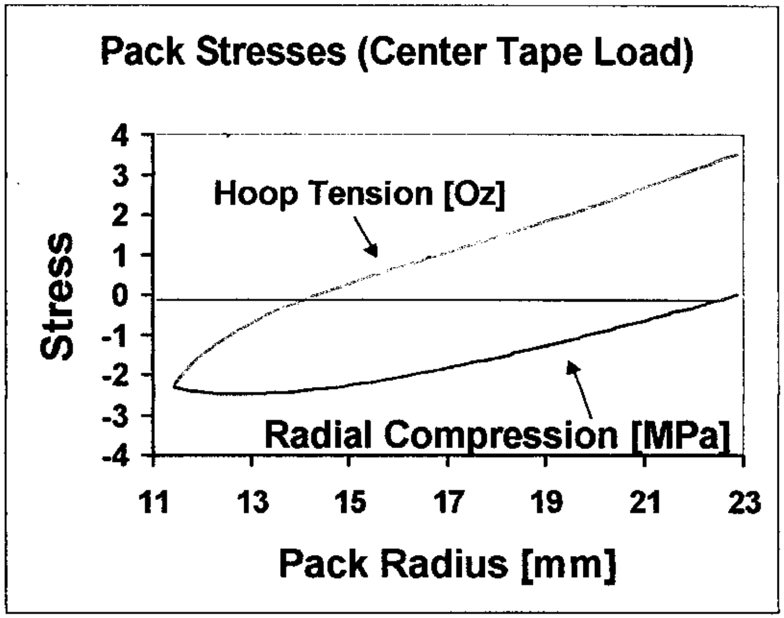


Figure 5. Results of Altmann spreadsheet modeling for tension in ounces and radial stress in MPa (for convenient use of the same numerical scale on the y-axis for stresses). With our 9840 plastic hubs, the hoop tension crosses over below zero to "compressive tension" at inner pack. The sum of hoop stress and radial stress (T + P in the same units), then has a fixed point somewhere towards the pack ID. These curves match those from Wickert numerical modeling.

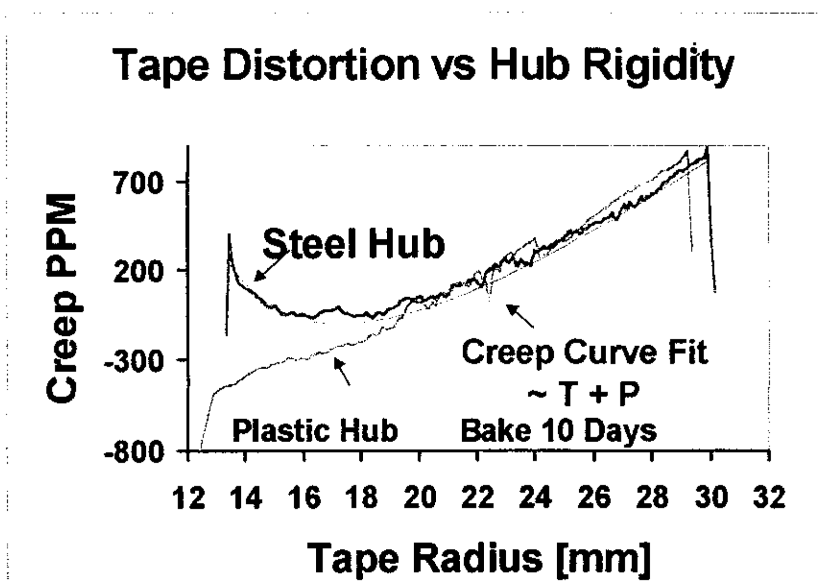


Figure 6: Comparison of 9.6-day creep for a "soft plastic" 9840 spool hub versus a stiff steel hub for end loaded 9840 tapes. These detailed and complete profiles cover the entire tape from hub-to-hub. The plastic hub tape creep curve is profile is nearly "linear downhill"

Brief Summary of Collapse to a Schwarzschild Black Hole

Dave Peterson, 11/13/17 (note for Cosmology+).

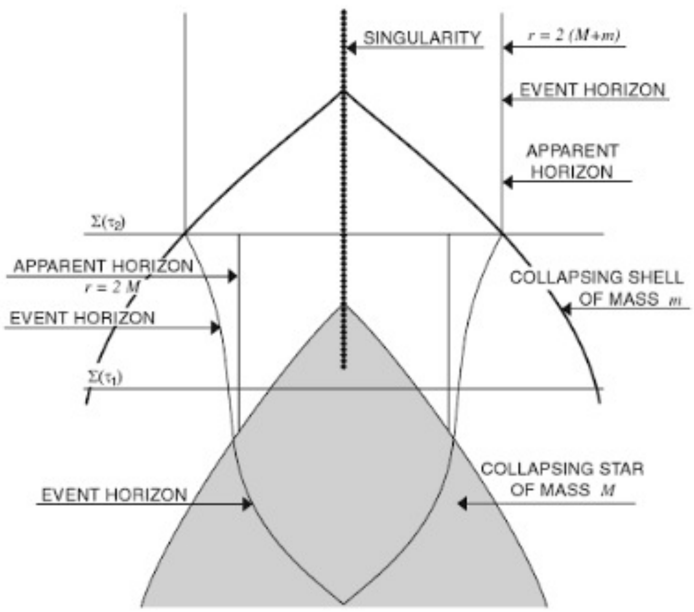


Figure 1. Radius of a collapsing massive ball with proper time on the vertical axis. An additional spherical mass shell then falls toward the black hole horizon.

The study of Black Holes (BH) presents many counter-intuitive challenges to nearly all of us. Part of the problem is the strong difference in view between using coordinate time (with respect to distant observers) or proper time (on the clock that moves along with an in-falling body or shell). In proper time, a body approaching a black hole simply and smoothly falls through the Schwarzschild horizon right on to a central singularity where it is utterly destroyed. This smooth fall is shown in Figure 1 in the boundary of the grey region; the vertical axis being proper time. But seen by a distant observer, a falling body or collapsing star freezes at the horizon leading to the "frozen star" view of black holes. Weinberg (1972) says: "The collapse to the Schwarzschild radius appears to an outside observer to take an infinite time, and the collapse to R = 0 is utterly unobservable from outside" [still true in 2017]. Its light doesn't suddenly disappear but does fade out of sight due to gravitational red-shifting. A century after Schwarzschild, some finer points and the interior behavior are still the subject of debate. The fall and fade of collapse can be seen in web movies [from CU]. As for mathematical descriptions, since Einstein's general relativity theory (GRT) is required, calculations can be challenging.

One of the first black hole collapse publications was by Oppenheimer and Snyder (OS) in 1939, "On Continued Gravitational Contraction" [1]. They considered a spherically symmetric finite radius ball of pressure-less dust collapsing freely in its own finite proper time (time in the frame of the moving body). Collapse models may treat a collapsing interior metric first and then dovetail it to an appropriate exterior model second. The initial ball free-falls in on itself because it has enough mass (say like 10 suns for example) to clearly exceed that of neutron stars and overcome any opposing degeneracy pressure. The profile of the equation of motion $r = r(\tau)$ for a ball of uniform density will show collapse just like that for the end of a closed universe model (using the usual FLRW metric for the cosmos). The fall towards its "final crunch" is a "cycloid" graph even for the Newtonian case r = r(t). A big star will collapse into a black hole with the formation of a one-way membrane called the event horizon (EH) that is then followed by a crunch to a central singularity S where density, curvature, and gravitational field become infinite [Figure 1]. We can also talk about a similar "apparent" horizon at each instant that first forms when a stellar surface crosses it's Schwarzschild "gravitational radius, $r \le r_s$ " -- but only for the OS type of collapse. More about that later.

The final outcome of collapse is viewed from an external observer "O" far away at large times. The collapse horizon is defined as a radial shell location of closest approach at which radially directed red-shifted light is just barely able to escape to infinity. What we call the "event" horizon is an "absolute" invariant surface defined with respect to asymptotically flat spacetime ideally (but not practically) at infinite distance and infinite future time labeled as "future null infinity", $\mathcal{I}^+$, and commonly called "scri " for script i. It is often pictured on paper as a future end-line of light rays transformed down from infinity to just a new convenient angle-number distance by using an arc-tangent function, $\pi/2 = \tan^{-1}(\infty)$. Infinity then becomes just an inch or two on a drawing.

This horizon occurs at the familiar Schwarzschild radius $r_s = 2MG/c^2 = 2m$ (where "m" means MG/c^2 with a standard unit convention of G = c = 1). At that radius from mass center, the gravitational field is $|g|=GM/r_s^2 = c^4/4GM = 1/4m$, and the Gaussian curvature is $K = 1/r_s^2 = 1/(2m)^2$. The Schwarzschild coordinate radius from a mass is defined so that circumference ($C=2\pi r$) and area ($A=4\pi r^2$) are the same formulas as for usual Euclidean geometry. If we conveniently assume the same for volume (not strictly true), then density = mass/volume = $\rho = 3c^6/32\pi G^3M^2 = 3/(32\pi\ m^2)$ --e.g., the Milky Way mega-Black Hole Sgr A* density is about the same as that of water).

In the Oppenheimer "toy model" case, a series of consecutive mass shells making up the uniform ball all fall to center at the same proper time τ [2] --just like the Newtonian gravity case for t. But real stars have strongest density at their centers (and have pressure and radiation) so that the central shells converge first. Then the event horizon (EH) originates first at the center [as in Figure 1] and expands outwards over increasing proper time as more shells "fall in" (the horizon is dynamical). This means that a far observer will first see the center become faint (if matter were ideally transparent) and then spread out to the limb over a time similar to that of light traveling a distance r_s. If the center were to momentarily become a black hole, then it would have infinite density, curvature and gravity (i.e., m or $r_s \rightarrow 0$ in the equations above). But the central density profiles in real stars aren't too strongly peaked, so "zero" might just mean "in the neighborhood of center zero."

From our far perspective, coordinate time, t, is really frozen close to the horizon ($g_{oo} \rightarrow 0$ and $g_{rr} \rightarrow \infty$), and we will never see a particle penetrate the horizon. Sparse falling matter or inwards directed light will be seen to accumulate just outside the EH. Once a black hole actually exists (meaning that somehow its mass is now essentially all interior to $r = r_s$), consider the special case that another mass shell is still infalling from the future (say M_{new} = 20% M_o). There should always be some additional matter falling in. Since EH is a far view, it anticipates this new infall and smoothly begins to widen [5] [and see Figure 1]. There are equations describing this continuous expansion of EH before complete merging of masses, and they act to smooth out the AH discontinuities of Figure

1 . After the infall, the new EH and AH is at $r_s = 1.2\ r_{so}$ and has expanded to engulf the new shell (the horizon crosses it!). This anticipation is called "teleological" (it incorporates future history as discussed in Kip Thorne's books [3]). The other convention called the "apparent horizon" (AH) only expands **at** the merging – a "fait accompli" view. Beyond that, the two conventions are the same.

Stephan Hawking considered the "apparent horizon" with some contempt and uses the phrase "absolute horizon" for our event horizon (>1970). The absolute horizon is smoothly increasing with new matter, is continuous and teleological, and looks at signals that "can" eventually just make it to the distant universe. The apparent horizon can be thought of as the boundary of black hole for light at this instant. Since it is not defined with respect to "future null infinity", it may not be invariant (and we strongly care about this). The AH separates light rays that are trapped inside a black hole from those that can move away from it, and the AH radius is always ≤ EH. Since the mass shells inside the horizon continue to fall to the center singularity over proper time, there is also an "inner AH" that begins with the formation of a horizon as a trapped surface boundary compelled to infall. This interior trapped boundary then falls to the singularity. Interior fall may also be described loosely in interior coordinate time which isn't too different from interior proper time and converges with it at the center [2]. Again, rather than coordinate time used by distant observers, proper time is carried along in the frame of moving bodies. Of course for the exterior case, coordinate time is drastically different from smooth fall in proper time.

OBSERVATION

Knowledge from observation comes mainly from deducing collapsed mass of unseen objects that are part of binary systems and matching them to theory. Distinguishing between a hard-surfaced neutron star and a black-hole may be aided by x-rays from accreting matter. But for hints about size, a current primary hope is the Event-Horizon-Telescope (EVT). This uses combined data from whole-earth Very Long Baseline Interferometry (VLBI) that may succeed in imaging the mega-black hole in the center of our galaxy [6].

Otherwise, we merely have un-practical heuristic aids for conceptual understanding .

Two views: We suppose that a far observer "O" at rest has a clock and records measurements over time. Since it is in flat space without gravitational curvature, coordinate time is the same as proper time.

One: Suppose we have a uniform lattice of numerous little bright blue light LED emitters filling the matter and space of a collapsing system. A far-observer "O" looks for and records the most redshifted photons it can see versus time on its own clock (pick a particular very weak frequency for the light energy, $E_{threshold}$). Since we only care about gravity and rays of light, we may assume that matter is ideally transparent. In the beginning of collapse, there may be little red-shifting because no concentrated mass or black hole has yet been formed. Then, threshold photons begin to be seen beginning first near the center of collapse. We are looking for faintness, and the arrival rate of faint photons increases with black-hole area (which increases up to the time of the creation of the final event horizon, EH, and then becomes a some-what constant weak rate.

Two: Outer surface of Collapse: [Misner-Thorne-Wheeler Text p. 847] says: "Place an astrophysicist on the surface of a collapsing star, and have him send a series of uniformly spaced signals to a distant astronomer at rest." In a time approach to the EH, the spacings will widen and the light frequency will red shift. The net result is that the luminosity of the signals decay exponentially with time and weakens quickly. A confusion factor is that some of the light will come not from just outside the horizon but rather from photon orbits at r = 3m with redshift z ~ 2. Worse than that, if we could see the light as representing a visual radius for the black hole, it would be at $3\sqrt{3}$ m ~ 5m, which is broader than the "photon sphere" at 3m that surrounds the hole (which itself is supposed to be at 2m). There is a lot of distortion, and decipherment requires that appropriate math has to be very carefully worked out.

The singularity: For the case of the OS homogeneous collapsing ball, the central singularity S begins after the collapsing mass falls inside its calculated Schwarzschild radius (r_s=2m), and the mass has completed its journey in proper time to the center. Since all the homogeneous mass arrives at the same time, there is no ambiguity about when S begins. "Cosmic Censorship" says that this singularity will (almost always) be shielded from view by the event horizon. For more centrally concentrated initial densities in the collapsing ball, S will begin before all the mass arrives at center [as in Figure 1]. Is S physical? With a little bit of rotation of the black hole, it may begin to broaden into a "ring-singularity". And with quantum-gravity, there might be a limit or even a bounce to the compactness of the center [4]. An interior Schwarzschild solution is known for the OS case of a constant density star, and it dovetails to the exterior solution. Although this ideal interior Schwarzschild solution is simple, real physics might be different. And we should note there is no really viable interior Kerr solution for more common case of strongly rotating black holes.

References:

[CU] Andrew Hamiltonian, CU., Collapse to a Black Hole, GIF Movies, casa.colorado.edu/~ajsh/collapse.html

[1] Oppenheimer and Snyder (OS) 1939, "On Continued Gravitational Contraction," [Phys Rev 56p455 or at: www.weylmann.com/oppenheimer2.pdf]. And also, Datt, S. Über eine Klasse von Lösungen der Gravitationsgleichungen der Relativität. Z. Phys. 1938, 108, 314–321 (independent and a little before Oppenheimer).

[2] Reflections on Relativity, web book, http://mathpages.com/rr/rrtoc.htm at 7.2, The Formation and Growth of Black Holes] http://www.mathpages.com/rr/s7-02/7-02.htm and also see: [http://mathpages.com/rr/s6-04/6-04.htm], MAA is The Mathematical Association of America.

[3] Kip S. Thorne, Black Holes & Time Warps, Norton, 1994, p 289+, and "Absolute Horizon" on p. 415 .

[4] Daniele Malafarina, Classical Collapse to Black Holes and Quantum Bounce: A Review, Universe, 2017 http://www.mdpi.com/2218-1997/3/2/48/htm

[5] Robert F. Penna, "Apparent motion of a spherical shell collapsing onto a black hole," in arXiv.org, type in number 1112.3638 gr-qc 2011 [closed universe Friedman cycloid fall arc-angle η gives shell $r = (r_i/2)(1+\cos\eta)$].

[6] https://www.nature.com/news/how-to-hunt-for-a-black-hole-with-a-telescope-the-size-of-earth-1.21693

Figure 1 from: Einstein's Field Equations and Their Physical Implications: Selected Essays in Honour of Jürgen Ehlers (Lecture Notes in Physics) Mar 15, 2000, Springer, edited by Bernd G. Schmidt .

[7] L. Rezzolla, "An Introduction to Stellar Collapse to Black Holes," - Max Planck ...
www.aei.mpg.de/~rezzolla/lnotes/IMPRS_2010/lect_V.pdf 2010

Appendix: The Early Event Horizon for r ≥ 2m

The "concave-up" early time portion of the event horizon, EH(τ), at the bottom of Figure 1 remains counter-intuitive. One approach to a derivation of this is assisted by the using "cycloid" coordinates, η = cycloid time (or cycloid parameter $0 \le \eta \le \pi$) and chi χ = hyperpolar angle [Penna 5] [Rezzolla 7].

{ The term **"Early EH"** means from proper time tau: $\tau_o = 0$ at bottom of Figure 1 up to τ_{2M} =creation of the apparent horizon, AH. Dust Ball radius falls from an initial R_o down to R_{AH}=2M. And cycloid time $\eta_o = 0$ up to η_{2M} .

For a preliminary perspective on geometry, first begin with the elementary distance metric for a **two**-dimensional spherical surface $\mathbf{S^2}$ (θ,ϕ) with radius **a** (e.g., a basketball with a ~12cm). This can be written as:

$\mathbf{d\sigma^2} = a^2 d\Omega^2 = a^2(d\theta^2+\sin^2\theta d\phi^2) = d\rho^2/(1-\rho^2/a^2) + \rho^2 d\phi^2$,

where ρ = radius from a vertical axis line through the sphere, and $d\Omega^2$ is an increment of spherical "solid angle." This latter metric is called a "Schwarzschild" form because circumferences are simply still $C = 2\pi\rho$. On paper, draw a circle and pick some upper angle θ from north and draw its ray from center to the point, p, on the circle arc. Then draw a horizontal line from the y-axis to p. This distance is ρ, and we will add an increment $d\rho$ onto it. Sketching out an incremental arc of $ad\theta$ at p and a base of $d\rho$ as a tiny triangle, we see that surface arc-length $ad\theta = d\rho/\cos\theta = d\rho/\{1-(\rho^2/a)^2\}^{½}$.
The Gaussian curvature of the sphere is $k = 1/a^2$, so the terms under the radical may be written as $(1- k\rho^2)$. The realization $\mathbf{d\rho = a\cos\theta\, d\theta}$ is Key because it means that $\mathbf{\rho = a\sin\theta}$ for a transformation between the two metric forms above (diffeomorhism).

Next consider a closed cosmology Friedmann metric (or "Friedmann–Lemaître–Robertson–Walker, FLRW", to give appropriate credits) metric cosmology as a time-size-changing 3-sphere $\mathbf{S^3}$ (χ,θ,ϕ) where we've added a third angle χ and replace $ad\theta$ for S^2 with $ad\chi$ (and $r = \sin\chi$) to get a spatial metric:

$d\sigma^2 = a^2(t)[dr^2/(1-\kappa r^2) + r^2 d\Omega^2]$

The collapse of a dust sphere to a black hole in the simplified OS collapse versus proper time is similar to this closed collapsing Friedmann (FLRW) cosmology.

Expressing this metric directly using the new hyperpolar angle χ gives:

$\mathbf{d\sigma^2 = a(t)^2[d\chi^2 + \sin^2\chi d\Omega^2]}$, and $\mathbf{ds^2 = -d\tau^2 + d\sigma^2}$ (c = 1 is understood).

For a fixed θ and ϕ angle, we have $d\sigma = a(t)d\chi$, or $d\chi = d\sigma/a(t)$.

[With $r = \sin\chi/\sqrt{\ }\kappa$, the previous denominator $(1- \kappa r^2)$ became $1-\sin^2\chi = \cos^2\chi$ to cancel that term in the numerator, $dr^2 = \cos^2\chi\, d\chi^2$. This metric is also good for the interior of the OS collapsing ball].

The time-varying $\mathbf{S^3}$ universe has no intrinsic definite length reference {Text MTW says that there is "no fiducial epoch"}. So a(t) and/or curvature may be arbitrarily rescaled away from some fixed meaning such as the basketball case (e.g., by $\kappa = k/a^2(t)$).

One <u>difference</u> between this FLRW universe metric and that for OS is that outside the dust ball R= R_o, the metric switches to the usual free-space Schwarzschild metric [7] for a central gravitational body, and R_o <u>is</u> a defined initial reference length.

At the boundary R = R_o, this **metric has to match** the outer Schwarzschild metric based on its coordinate r as usual circumference about a mass source: C = $2\pi R_o$, which also equals $\int (g_{\phi\phi})^{1/2}\, d\phi = 2\pi a(\tau_o)\sin\chi_o$, or $R_o=a(\tau_o)\sin\chi_o$, an initial value for χ at wide R_o.

<u>Proper interior elapsed time</u> for the ball is given by the integral from R to r:
$\tau = \int dR/[2m(1/r - 1/R)]^{1/2} = \{(R/2m)\}\cdot \{(r(R\text{-}r)]^{1/2} \text{-}(R/2)\sin^{-1}(2r/R -1)\} - (R/2m)^{1/2}\cdot \{(R(R\text{-}R)]^{1/2} - (R/2)\sin^{-1}(2R/R -1)\} =$
$\tau = (R/2m)^{1/2}\, [\, (r(R\text{-}r))^{1/2} - (R/2)\sin^{-1}(2r/R -1)+(R/2)(\pi/2)\,]$

Simplify the answer for the time to fall with new angle variables:
Let the factor (2r/R -1) = **cos η** , so **<u>$r(\eta) = R(1+\cos\eta)/2$.</u>** Substituting this r(η) above to get: τ = $\boldsymbol{\tau(\eta) = (R_o^3/2M)^{1/2}(\eta + \sin\eta)/2 = a_m(\eta + \sin\eta)/2}$

{We used $\sin^{-1}[\cos\eta] = \pi/2 - \eta$; and a_m means max a}. [ref. 7] and see "MathPages" [2]. Switch notation R to r becomes R_o to R [So $\chi_o = \sin^{-1}(R_o/a(t) = R_o/(R_o^3/2M)^{1/2} = 1/(R_o/2M)^{1/2}$)]. For the initial time = 0 at the bottom of the Figure 1, we have η =0 (i.e., 0+sin0 = 0 = τ) and R = $R_o(1+1)/2 = R_o$.
R(η) and τ(η) are called <u>cycloid relations</u>.

[In contrast to proper time τ , outside Schwarzschild coordinate time t = $t(R_o,\eta)$ about a central mass is a very complicated expression not shown here, see [5] eqn. 6].

Notice that $d\tau(\eta) = a_m(1+\cos\eta)d\eta/2 = d\eta\, a_m r/R$ or $=d\eta\, a_m R/R_o = d\eta\, a(\tau)$.
In cosmology, the present separation of stars: $D_{now} = \int c\,d\tau/a(\tau) = \int c\, d\eta$
from the time of light emission t_e to the present time now = t_o.
Alternatively, we could have said: $ds^2=0=-c^2dt^2+a^2(t)d\chi^2$ means $d\chi = cdt/a(t)$.

The full FLRW metric is: $ds^2 = \boldsymbol{a(\tau)^2\, [- d\eta^2 + d\chi^2 + \sin^2\chi d\Omega^2]}$.

The proper time τ = $\tau(R_o,\eta)$ from η =0 to max value η = π gives the finite total proper <u>time of fall</u> from R_o to the singularity S at R=0 as: $\tau_{fall} = (\pi/2)\,(R_o^3/2M)^{1/2}$.
For a uniform density ball, $R^3/m(R)$ is a constant; so proper time is a constant – <u>all</u> shells will fall to zero at the same time! But this is true only for the OS constant density model.

Now from the matching condition $R_o = a_{max}(\tau)\sin\chi_o = (R_o^3/2M)^{1/2}\sin\chi_o$. So the beginning of fall is at η = 0 and $\chi_o = \sin^{-1}(2M/R_o)^{1/2} = \sin^{-1}(R_o/a_m)$ [7].

The time of fall from η =0 to the 2M event horizon is $\tau_{2M} = (R_o^3/2M)^{1/2}\,(\eta_{2M} + \sin\eta_{2M})/2$.
So collapse has landmarks : $\eta = 0 \leq \eta_{2M} \leq \pi$. And $\chi = \chi_o \leq \pi$ (note that ref 7 eqn 49 forgot the /2 – very important).
For the early event horizon, EH, Rezzola says "study the trajectory of the outermost outgoing photon that was not able to reach null infinity," and at each instant during collapse the last outgoing photon that will be sent and reach null infinity [7].

Outgoing photons have $ds^2=0$, so from the angular metric we have $d\chi/d\tau = \pm 1/a(\tau)$. Their trajectory obeys $d\chi/d\eta = \pm 1$ (chi is spatial, eta is temporal). The "place and time of emission" obey
$\chi = \chi_e \pm (\eta - \eta_e)$. [the slope +1 gives a $-(\eta - \eta_e)$]. "A swarm of outgoing photons will be **trapped** if their proper area will not grow in time," **$dA/d\eta \leq 0$** where Area $A = \int (g_{\theta\theta}\, g_{\phi\phi})^{½}\, d\theta\, d\phi = 4\pi\, a^2(t) \sin^2\chi$ (so un-trapped = free uses >, and free means can go to infinity).

Now $a(\eta(\tau)) = (a_m/2)(1+\cos\eta)$, contributes an eta factor of $(1+\cos\eta)^2$. But the chi curves include eta also as $\chi = \chi_e - (\eta - \eta_e)$. So η derivatives have to include that as well using a product rule (the simple math hides a lot of physics).

Rezzolla [7] claims that the net result of trapped area is **$\eta_e \geq \pi - 2\chi_e$** (a region of the χ,η plane, $\chi_e \geq \pi/2 - \eta_e$). [And that is true from comparative graphs. The factor of 2 comes from a term $2\sin\chi\cos\chi = \sin 2\chi$].

The apparent horizon AH is defined as the outermost trapped region trying to emit within the star.
$\eta_{ah} = \pi - 2\chi_o = 2\cos^{-1}(2M/R_o)^{½}$ [using the previous $\chi_o = \sin^{-1}(2M/R_o)^{1/2}$ and $\sin\chi_o = \cos(\pi/2 - \chi_o)$].

Notice that if we set 2M = 1(some distance unit), since $\cos(2x) = 2\cos^2 x - 1$, where $x = 2\mathrm{acos}(1/R_o)^{½}$, then radius $R = (R_o/2)(1+\cos[2\,\mathrm{acos}(1/R_o)^{½}]) = (R_o/2)(1 + 2/R_o - 1) =$ **1.0** also. So after AH, Radius **R_{AH} = 2M stays constant!**

Examples of values at R_o=2M, $\eta_{AH}=0$, $\chi_o = \pi/2$, $a_m=1.0$, $\tau = 0$ [Google Sheet].
For R_o= 4M (double wide) $\eta_{AH} = \pi/2$, $\chi_o = \pi/4$, $a_m = 2.83$, $\tau = 3.64$. But collapse to R =0 gives $2.83(\pi+0)/2 = 4.44$ units > 3.64 (since there is more collapse after R=2M).

Formation of the Event Horizon, EH: $R_{EH} \geq R_{AH}$, but AH only forms at R = 2M and above that we have $\chi_{EH} = \chi_{AH}$ (equality holds) when $\eta = \eta_{AH}$. The "worldline for the event horizon is given by": $\chi_{EH} = \chi_o + (\eta - \eta_{AH})$ for $\eta \leq \eta_{AH}$.
Recall $R = 0.5\, a_m(1+\cos\eta)$. But, for circumferential radial coordinates, R = Cir/2π or R = $(\text{Area}/4\pi)^{½}$, we have a θ and ϕ metric coefficient also of $\sin\chi$. So,
$R_{EH} = 0.5\, a_m(1+\cos\eta) \sin(\chi_{EH} = \chi_o + \eta - \eta_{AH})$.

Now, $\chi_o - \eta_{AH}$ is **negative**! So as η increases from 0 at $R=R_o$, there is some positive eta at which $\chi_{EH} = 0$! That means that the event horizon begins at the middle of the collapsing dust ball and then widens. $\chi_{EH}=0$ for $R_o/2M = 2$ with $\eta = +0.8 < 3.64$. Or for $R_o/2M=3$, $\eta = +1.3 < 7.41$. So these beginnings of EH lie between $R = R_o$ and R = 2M.

A spreadsheet plot duplicates the shapes in Figure 1. Column of eta's from 0 at 4M to 4.4 at R=0. The EH begins near $\eta \sim 0.8$ $\tau \sim 2.15$ curves upward to the intersection with the AH (R=2M) near $\eta \sim 1.5$, $\tau \sim 3.53$ (should be 3.64).
We might say that the cause of an interior event horizon is due to viewing from Schwarzschild space far away where radial coordinates are compatible with circle circumferences C = 2πR.

Boulder Cosmology and Modern Physics Group: Questions and Comments

{Boulder Library Meetings} Dave Peterson, last 11/12/19

SAMPLE TOPICS:

Comments and Addenda to Group Selected Book Readings
Observational Cosmology, Stephen Serjeant
An Introduction to Modern Cosmology, Andrew Liddle
NOETHER'S THEOREM
Questions: Gravitational Energy (?), Friedman Equations,
Time curvature and Newton's gravity
Previous Book: Basic Concepts in Physics, Masud Chaichian
Entanglement Swapping.

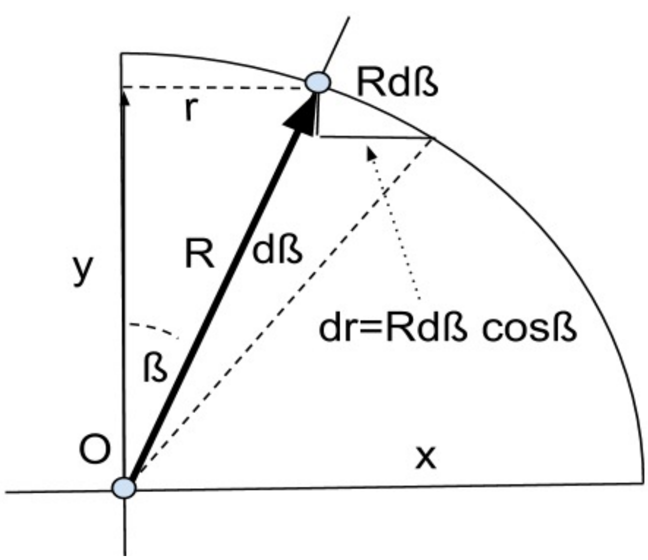
Addenda to: Observational Cosmology

TextBook by Stephen Serjeant , 8/23/19 -9/1/19

For Meeting on 9/16/19: Some comments and special additions that perhaps "should" have been somewhere in our new Book for Cosmology and might answer some questions. Chapter Two is longer and harder than Chapter One.

The Space Metric for a Basketball:

Serjeant just states a metric for a spherical space S^3 in eqn 1.6 (and uses it in 1.37). Where does his $\mathbf{dr^2/[1\text{-}kr^2]}$ term come from? It helps to first have a clear explanation for a simplest case like the surface of a basketball {or spherical shell, S^2 } with polar angle θ , longitude angle ϕ and radius R. That is easily done by examining a curve portion like that shown in the **Figure** below. Pick any longitude, say ϕ = 0, and only look at circular arcs in θ. Pick a point on the sphere and let r be the "radius" to that point from a y-axis.



The usual differential angle space metric here is $(d\ell)^2 = R^2(d\theta^2 + \sin^2\theta\, d\phi^2)$, so an element of θ arc {or β in the figure} has familiar length $d\ell = Rd\theta$, and the element of length around a latitude is $R\sin\theta d\phi = rd\phi$ where $r = R \sin\theta$. Examine a tiny differential triangle having acute angle θ again, hypotenuse $Rd\theta$, altitude dy, base $dx=dr=Rd\theta \cos\theta$. Now $\cos \theta = y/R$ where $y = \sqrt{}(R^2 - r^2)$, so $Rd\theta = dr/\cos\theta$;
and $\cos\theta = (1-\sin^2\theta)^{1/2} = \sqrt{}\ (1- r^2/R^2)$.

So, $(d\ell)^2 = (Rd\theta)^2 + (r\, d\phi)^2 = \mathbf{[dr^2 /(1- r^2/R^2)]} + r^2 d\phi^2$.

And, the curvature of a sphere is "k"= $+1/R^2$. ……
…And then we play games with cosmological scale and scale factors and address "three-sphere" metrics embedded say in 4-dimensional Euclidean space. We could now discuss S^3 using three angles: θ, ϕ and a new "hyperpolar angle" chi, χ, that we can't easily picture). Then, it will be the new $(Rd\chi)^2$ term that will be equal to $dr^2/[1- r^2/R^2]$ in equation 1.6.

A touch of History for expanding cosmology:

Einstein proposed his static universe cosmology in 1917 using λ as a term counteracting gravity (at that time, our milky way was the whole universe – so the idea of a homogeneous isotropic universe was inspired – or convenient). de Sitter immediately published his own model universe without matter and using only λ. Then in 1922 Friedman considered a dynamic radius of curvature R = R(time) – his new universe could expand or even oscillate. In 1927, Lemaitre also proposed an expanding universe. Einstein rejected both proposals. In 1930, Eddington stated that Einstein's 1917 static world solution was unstable and might easily expand or contract. So, in 1931 Einstein finally agreed that the model of the universe should be a dynamic one like Friedman's and abandoned the cosmological constant.

{See "Einstein's conversion" at https://arxiv.org/pdf/1311.2763.pdf }.

Cosmological Distance in Chapter One:

Brief Summary: We seem to have **seven** (or more **) types of distance!**
One is just ruler or metric distance d_p between masses ("proper" distance separation at the same time – any time, not limited to light emission and absorption). Or, we could say, "Cosmological proper distance" between two points measured along a path defined at any constant cosmological time ($d_p = a(t)\, \Delta R$). In Chapter two, Sergeant uses $r_p = \int cdt/R(t)$ as proper distance. Eqn. 2.12.

Then three deduced light distances. Let "then" be a time when a galaxy emitted light and "now" when we receive it. Emit distance d_e is 'emit to receive' distance both at time = "then" = "d_p then." Look-back time or "light travel" distance $d_{LT} = c\Delta t$ from there and then to here and now. **Comoving** distance $d_C = \int cdt/a(t) = \int^z_o cdz/H(z)$ includes the expansion of space from "there and then" to "here and now" -- where the source and receiver are now, d_c is d_p "now" and so is also called d_{now} or d_o (i.e., when $a(t) = a_o = 1$).

Distances ordering is $\mathbf{d_{emit} < d_{LT} < d_{now}}$.

We also use $d_{hor} = d_{horizon} = \int cdz/H(z)$ from 0 to ∞ (from emit time ~ zero!).
The "Particle" (or cosmological or comoving or light) Horizon is the maximum distance from which light could have traveled to the observer over the age of the universe – the size of the observable universe.

Three observed distances: Angular diameter distance d_A= object diameter/$\Delta\theta$; "proper motion distance" from transverse speed $d_M = v_\perp / \Delta\omega$ where $\omega = \Delta\theta/\Delta t$ -- also coincides with $r_e = r_o - r_{emit}$ or "coordinate distance measure." And there is Luminosity distance d_L using observed light flux.
$d_A=a^2 d_L$ and $d_M = a d_L$ (and $d_A = a d_M$, $a \leq 1$), so ordered distances are $\mathbf{d_A < d_M < d_L}$.

For more, see "Misconceptions" at https://arxiv.org/pdf/astro-ph/0310808.pdf and http://astro.pas.rochester.edu/~aquillen/ast142/Lecture/cosmo.pdf

The **Text Equation 1.33** for Hubble ratio $H(z)/H_o = E(z)$ is important, is used, presents problems, and looks like it deviates from everything I've ever previously seen: {Such as Peebles' Cosmology pg. 100: $(H/H_o)^2=\Omega_{mo}(1+z)^3+\Omega_{ro}(1+z)^2 + \Omega_\Lambda \equiv$ "$E^2(z)$". Similarly, Misner,Thorne,Wheeler {**Gravitation**, "The Telephone Book"} eqn. 27.40 is nearly the same but with scale a instead of z.

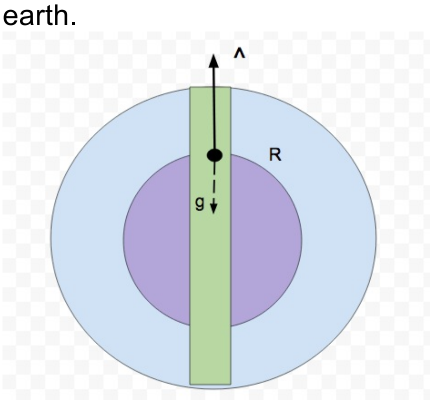
$(a\ \text{dot})^2/a^2 = -k/a^2 + \Lambda/3 + (8\pi/3)(\rho_{mo}\, a_o^3/a^3 + \rho_{ro}\, a_o^4/a^4) \}$.

The H/Ho = E(z) formula by Serjeant must work ok but is hard to "grok." {He set $\Omega_r = 0$ here and discarded curvature k}. He uses his equation in 1.34 and again in 1.44 &1.56 . {Bill Daniel has written out the algebra for the derivation of 1.33.} Without radiation, Sergeant's equation has limited range {Chela has commented on this}-- perhaps out to $z \leq 5$ -- which is adequate for Observational cosmology.

EdS The **"Einstein-de Sitter"** cosmological model of 1932 has only mass $\Omega_{m,o}=1$, and $\Lambda = 0$ (even though the de Sitter universe was all Λ). It has the great virtue of easy calculations in closed form (vs numerical integration otherwise) and works fairly well for 300<z<2. So it is good for homework exercises (like Ex.1.4, 1.5, eqn 1.45, Eqn 4.7 and for simple understandings). It was very **popular** for many years—even in 1980 when it was discovered that $k \simeq 0$. Many books now don't even mention it {…I don't like to discard history}. In section 2.7, the "particle horizon" for an EdS universe is $d_{proper\infty}=d_{hor} = 2c/H_o$.

The simplest Way to introduce Cosmic Inflation: (see Section 2.7- 2.8)

A thought problem for a cylindrical shaft filled with vacuum going all the way through the earth.



The accelerating expansion due to inflation can be related to the freshman physics problem of the motion of a ball falling through a long hole dug through the center of the earth. At the surface of the earth, the gravity is g_o (e.g., 9.8 m/s^2). At any other radius away from center, the mass of the earth that contributes to attraction is only the mass inside a spherical "Gaussian surface" at that radius, R. Near the center, that volume is tiny so that there is little force. As the body moves outwards, there is more and more attracting mass below the ball, so the restoring force increases and the body comes to a halt.

Force = F= - kR = mass·acc = m· d^2R/dt^2. The period of oscillation is found to be tau =τ = $\sqrt{}${3π/ρG} ≃ 1.4 hours (where average earth density is 5.52 g/cc) . The ball simply falls through the earth to the other side and then back again. Because of the negative sign; the solution is just simple harmonic motion like that of a spring with a restoring force – a **SINE** Wave.

Now switch to **Λ** and change signs on the spring constant! **-** → + .
Inflation with a huge cosmological constant and with p = -ρ would end up with a net negative -2p anti-source causing effectively a **repulsive gravity** which makes the universe `fall outwards.' Or, we might consider a spherical shell of `pebbles falling outwards.' This form has a repulsive force **F = +kR,** a similar but different differential equation. Every step away from the center of the earth sees more "mass" behind it with more and more repulsive force. Instead of sine-wave motion, the solution this time is a runaway **exponential expansion**! {a "little" difference is that inflation has no "center."}
[exercise: plug R = $R_o$sinωt and also R = ke^{+bt} into $d^2R/dt^2 = \pm kR$ to show that the signs work out right]. The inflation solution is:

$R(t) = ke^{\{+bt\}}$ where b = $\sqrt{}$ {8πG ρ /3}.

Two problems are, ``how does it start and how does it end?"

https://en.wikipedia.org/wiki/Inflation_(cosmology)
http://w.astro.berkeley.edu/~jcohn/inflation.html

The discussion of inflation in our book sections 2.7,2.8 is not easy to grasp with clarity.

Recall the two Friedman equations (1.7 &1.8): a first order one with a $(dR/dt)^2$ term and dynamic one of order two with a d^2R/dt^2 term. Given an intense scalar "inflaton" field with huge energy density V(ϕ), the dynamic equation produces an initial fast expansion that can be dampened by friction. Then, in the other equation, this expansion quickly makes any curvature contribution negligible ($k/a^2 \rightarrow 0$, p.56 eqn.1.7, 2.22,2.24) leaving a "possible" Λ and a residual scalar potential field V(ϕ) which can be considered nearly constant due to a "slow roll" nearly flat potential. I'll just lump these together into some new huge effective Λ (not our "traditional" or current cosmic constant Λ). A resulting $(dR/dt)^2 \sim \Lambda c^2R^2/3$ has a solution $R = R_o e^{\sqrt{(\Lambda c2/3)}t} = R_o e^{Ht}$ {rapid **exponential** growth! – like the repulsive gravity above}.

Sergeant avoids most of this commonplace simplicity and just ends up saying **$H^2 \propto V(\phi)$** {eqn. 2.24, which amounts very roughly to the simple math above} with no further discussion -- as if you should know what it means! (This equation is similar to the old de Sitter equation on pg. 36).

Note that there are so many different versions of inflation theory that it might not be falsifiable (possibly meaning "beyond science").

Planck Mass, m_{Planck}, using h, c, and G : Dimensionless constants were suggested in 1899 **before** the Black Body radiation paper of 1900 that introduced what was later called Planck's constant, h (Mike and I are still not sure how). {Ref: M. Planck. Naturlische Masseinheiten. Der Koniglich Preussischen Akademie Der Wissenschaften, p. 479, 1899} (m_{Planck} is used in our book, Section 2.7). The fields ϕ used in Inflation are near this mass energy ! (see answers 2.7 p. 295).

The scale invariant power spectrum (p. 63) with equal energy per octave can also be called 1/ f noise or Pink noise, and has a "random fractal structure."

The Speed of Sound, c_s, in the Universe at the time of Recombination, (CMB, z ~ 1000): … is a sizeable fraction of the speed of light!

For a photons-only (very early) universe without mass, $c_s=\sqrt{(p/\rho)} =\sqrt{(c^2/3)} = c/\sqrt{3} \simeq 0.58c$. But after the $\Omega_m \sim \Omega_r$ equality near z ~ 24,000, the inertia of matter begins to alter and reduce this speed. "Acoustic Peaks" Page 73 says that the speed of sound relative to the speed of light is $\beta = c_s/c = (3 + 2.25\Omega_b/\Omega_r)^{-½}$, so we need to know the baryon to radiation ratio.

Eqn 1.15 is $\Omega_r=8\pi G\rho_r/3H^2$. Then $\Omega_b/\Omega_r = \rho_b/\rho_r = \rho_{bo}/a^3 \ / \rho_{ro}/a^4 = a(\Omega_{bo}/\Omega_{ro})$ now. At present, the Ω fractions for "matter" (in this case being "dark matter"), baryons and radiation with h ~ 0.7 is roughly
$(m_o, b_o, r_o) \sim (0.26, 0.043, \sim 2\times10^{-5})$ or $\Omega_{bo}/\Omega_{ro} \sim 2150$ – highly matter dominated!

Then z ~ 1000 says that temperature at recombination is near 3000K which drops to the present TBB ~ 3K. Then, (a~ 0.0009)x(2150) ~ 1.98, so $\beta \sim$ 0.45c. Exercise 2.9 uses $\beta \sim 0.58$ – OK, but not exactly right. $c_s/c=1/\sqrt{3}$ is a commonplace conventional reference.

After the CMB, light pressure no longer counts and $c_s \rightarrow (4c^2\rho_r/9\rho_m)^{½}$. Temperatures of radiation and matter become nearly the same.

2.16 "The polarization of the CMB" "The detection of B-mode polarized clustering would be terribly exciting…" (p.80) .. and, an announcement of such a discovery was made in 2014. BUT: [Nature Jan 2015]: "A team of astronomers that last year reported evidence for gravitational waves from the early Universe has now withdrawn the claim. A joint analysis of data recorded by the team's **BICEP2 telescope at the South Pole** and by the European spacecraft Planck has revealed that the signal can be entirely attributed to dust in the Milky Way rather than having a more ancient, cosmic origin. (Our Sergeant book came out in 2010)

Addenda to Observational Cosmology, Chapter 3.

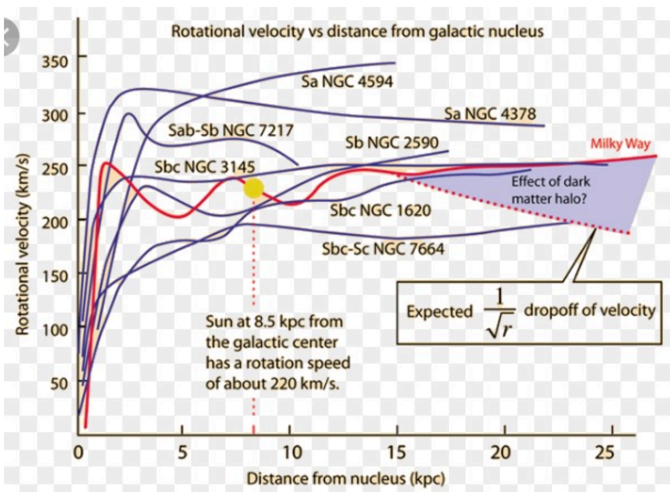
Dave 9/2/19 – 11/2/19

For meeting on 11/18/19:
Galaxy Rotation Curves, v(r).

The plot of rotational velocity versus radius from galaxy center shown in text rises more steeply than usual [Figure 3.1, p. 93]. Dark matter content varies from case to case, but "Most spiral galaxies show **flat rotation curves** out as far as we can trace them, even where no more stars are visible" (e.g., the Figure above). The rotation curve of our closest galaxy Andromeda M31(not shown above) is also very flat. The implication is that dark matter halos dominate and extend far beyond the visible disk of a galaxy.

The dark matter content of the Milky Way is about 90-95% which is higher than the 85% for the universe as a whole. The mass density for a general rotation curve can go roughly as $\rho(r)/\rho_o = 1/[1+r^2/r_c^2]$ where r_c is the radius of the galaxies central visible "core." So, for $r \gg r_c$, dark matter density $\rho \propto 1/r^2$. If $\rho(r)$ is spherically symmetric (and that varies too), then outermost velocities will be flat, $v(r) \sim v_{flat}$.

Note that estimates of the size of our Milky Way galaxy recently doubled (Gaia and HST data) now out to a radius near 130,000 light years and a total mass near 1.5 trillion suns (out to the outermost globular clusters). The number of visible stars in the MW is about 200 billion (so roughly 90% of the mass is dark matter halo). Our mass is now competitive with that of Andromeda M31. The extent of the DM halo may be ten times wider than the visible galaxy. https://arxiv.org/pdf/1804.11348.pdf



Chapter 3 The local universe

Figure Velocity curves for Spiral Galaxies {hyperphysics.phy-astr.gsu.edu}

Neutrino Equation 3.1 presents a strange and curious little puzzle: why sum up the neutrino masses and what is the meaning of the 93.5 eV value in the denominator. Neutrino masses are not well known but there is an experimental constraint on the sum of the masses of the electron, muon and tau neutrinos (perhaps $\Sigma m < 0.72$ eV). There is no individual identity since neutrinos transmute into each other over distance and time depending on their energy. The mysterious 93.5 eV reference value in the equation seems to be like the energy of an imaginary particle such that the same density of them as the neutrino triplet density would close the universe (a replacement for critical density). It is estimated that the number density of individual neutrinos now is roughly $330/cm^3$ at a temperature near 1.9 K.

Section 3.3 p 96: For the simple but unreal case of all galaxy mass concentrated near the central bulge, the discussion on **Tully-Fisher** in the text could be a bit more transparent. Extremes of Doppler shifts come from the visible edges of the galaxy. If we see a spiral galaxy "edge on", then one side is speeding towards us with velocity v and the other side away from us (so $\Delta v \sim 2v$). Gravity force = centrifugal force, $mMG/R^2 = mv^2/R = (m/R)(\Delta v/2)^2$; so $M \propto R(\Delta v)^2$. Also notice that kinetic energy is – ½ times gravitational potential energy: $KE = mv^2/2 = mMG/2R = (-\frac{1}{2})(-mMG/R)$. This is the simplest example of the virial theorem for gravitationally bound systems (p 99). Note that the virial theorem also applies to much smaller systems such as the ground state kinetic and potential energy of atoms and molecules [Ruedenberg].

Using the Virial Theorem, detected kinetic energies of stars in galaxies and of galaxies in galaxy clusters indicate what gravitational potential energy must be present. That in turn tells us the amount of unseen dark matter that must be present (e.g., page 115).

"kSZ" (bottom of page 101). "Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect," arXiv:1203.4219 (and Phys. Rev. Letters). "The Atacama Cosmology Telescope (ACT) performs the first statistical detection of the kinematic SZ effect." (This was in 2012, our book is dated 2010).

HISTORY: I was a bit appalled at the absence of human history in this book and in the development of the expanding universe and feel that some outside reading is desired to counterbalance that. For example, on Cepheid variable stars on page 105, it might have said: Henrieta Swan "Leavitt's discovery provided astronomers with the first 'standard candle' with which to measure the distance to faraway galaxies." This 1912 work was KEY to the great discoveries up to Hubble's law of 1929.
See Wikipedia: https://en.wikipedia.org/wiki/Cepheid_variable, and https://en.wikipedia.org/wiki/Henrietta_Swan_Leavitt,

HUBBLE H_O AND STANDARD CANDLE LIST: Red Giants as Standard Candles (bottom of pg 105) and the **dilemma** of two different values for Hubble H_o: You've all heard that recent local Hubble estimates indicate that the universe is growing 10% faster than indicated by analysis of the cosmic micro-wave background radiation (CMB). https://www.quantamagazine.org/cosmologists-debate-how-fast-the-universe-is-expanding-20190808/ : Recent research by Wendy Freedman says, "Using **tip-of-the-red-giant-branch stars**, they'd pegged the Hubble constant at 69.8 — notably short of SH0ES' 74.0 measurement using cepheids and H0LiCOW's 73.3 from quasars, and more than halfway to Planck's 67.4 prediction."

Time-delay cosmography (multiple images near gravitational lens distance measure) – **a new method** not listed on p 105 Chapter 3. 10/23/19 "New measurement of Hubble constant adds to cosmic mystery," … looked at light from extremely distant galaxies that is distorted and split into multiple images by the lensing effect of galaxies (and their associated dark matter) between the source and Earth. {The source galaxies are far away, but the lenses are near – like z~0.3 to 0.34}. By measuring the time delay for light to make its way by different routes through the foreground lens, the team could estimate the Hubble constant (H_o = 76.8 !, continually higher than Planck CMB). In 2017, the H0LICOW team published an estimate of 71.9, using the same method. There is now a 4.4σ tension between Planck and other local measures.(‼)
https://www.sciencedaily.com/releases/2019/10/191023150327.htm, and also
https://academic.oup.com/mnras/article/490/2/1743/5568378

Exercise 3.2 calculates the negative gravitational energy of a ball of matter (page 104). This alters gravitational mass and inertial mass to the same degree (absence of Nordtvedt effect). But, the gravitational mass is defined by the asymptotic Newtonian potential **at large distance** from the system—not close up. There is no real concept of close-up real gravitational potential energy in general relativity.

Pg 109 **Collision of Andromeda** with our Milky Way Galaxy, animation: **NEAT**!!
https://www.youtube.com/watch?v=fyQrdsTNuo0

109-112 The pictures of universe structure are very nice but are dated. There is a really nice 2019 picture of the huge local VOID on the web along with a great 4 minute movie animation of our **supercluster –Laniakea**.
https://www.universetoday.com/142923/meet-our-neighbour-the-local-void-gaze-into-it-puny-humans/

And Comments on Last Month:

Question 1**. Inflation** as $H^2 = 8\pi G\rho_\Lambda/3$ doesn't look like a "ball falling through Earth" spring type problem, F = +kR – so is it the same physics? Well, the Friedman "acceleration" equation is $a''/a = -4\pi G(\rho + 3P/c^2)/3$. But, for an "only Lambda" universe with constant density, conservation of energy implies that we also have $P_\Lambda = -\rho_\Lambda c^2$ (effective negative pressure!), so $a''/a = +8\pi G\rho_\Lambda/3$ – same as for H^2. And note that $(d/dt)(H) = (d/dt)(da/adt) = (a\,a'' - a'^2)/a^2 = a''/a - H^2 = 0$, so $a''/a = H^2$! So, yes, the equations mean the same thing.

FYI: **The (nearly) Latest Astrophysical Constants** can be seen at:
http://pdg.lbl.gov/2017/reviews/rpp2017-rev-astrophysical-constants.pdf
[For interest: The Astronomical constants sheet says $\Omega_m = \Omega_r$ at z~3400 and adds that the z where universe acceleration = 0 is only z_q ~ 0.65. Compare that to $\Omega_m = \Omega_\Lambda$ at scale factor $a = (0.31/0.69)^{1/3} = 0.76$, (z~0.31, perhaps 3.5 billion years ago – see on-line calculator http://www.astro.ucla.edu/~wright/CosmoCalc.html).]

Note: Friedman (1888-1925) was Russian with Cyrillic spelling, Фри́дман. Friedmann is a German form that is sometimes preferred in English (but not in spell-checker).

Addenda to Observational Cosmology, Chapter 4.

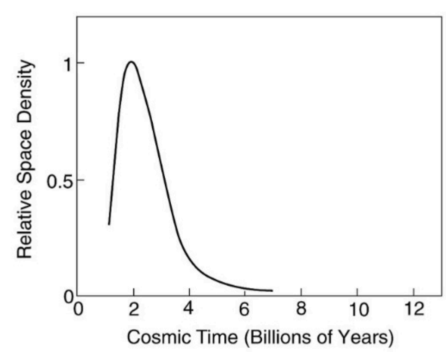
Dave 11/7/19 – 11/19/19

For meeting on December 16, 2019
{Beethoven's Birthday, 1770}

The distant optical universe

[Think about how many kinds of color filters one might use].

Figure 1: Quasar density evolution [ESO.org] →



Chapter 4 Page 121 has an important , *and opaque*, sentence that radio source counts steeper than an $S^{-5/2}$ power law counted against a steady-state ("SS") universe *(Fred Hoyle's 1948 pre-big-bang continuous-creation model – which had many believers).* { Translation: long ago and at big distances from us, source counts were unexpectedly high}.

In chapter 1, we learned that S = flux (energy seen per unit area of sky) obeyed $dN/dS \propto S^{-5/2} \propto r^5$ (where radius r is distance from us before present time). Now **quasars** and new stars and radio galaxies were dense for redshifts z >2 *but weak for z > 4* – meaning between about 4 to 1 billion years cosmic time after the big bang and peaking at 2 b years (shown in **Figure 1**). That is, there were big changes in the evolution of the universe over time. The history of the universe was decisively not steady state! Also, SS could not explain our black body CMB radiation, it was not due to old star-light scattering from galactic dust clouds. {Note for 2019: 203 quasars have now been discovered with $z > 6$ and defy currently accepted models. There is even a supermassive black hole at redshift $z = 7.54$. *"How did the first SMBHs grow so large so fast?"* Primordial BHs **[PBHs]** are being reconsidered as possibly major dark matter}.

It turns out that **Figure 4.17** on page 143 is another version of figure 1 above. Roughly, redshift z =1 means about 7.7 billion years ago (or "Ga" giga-annum before present). z =2 is 10.3 Ga and 6 is about 12.7 Ga—so the x-axis for redshift can be converted into time before us.

Similarly Figure **4.24** shows star formations peaking at z = 2, and this x-axis can also be converted to time in billions of years.

"Hot **dark matter** (HDM) candidates are relativistic particles, i.e., which move with velocities close to the speed of light, e.g. the neutrinos. Cold dark matter candidates are non-relativistic, i.e. slow moving particles." Although light weight, Axions (p.93) "are non-relativistic and therefore fall within the category of CDM." The unexpectedly high density of black holes suggests a re-thinking of that as cold DM.

Figure 4.1 on the Matter Power spectrum P(k) : Previously (on pages 60-64) we cared about the **clumpiness** of the little temperature or density variations of the CMB and noted a "Harrison-Zel'dovich" primordial $P(k) \propto k$ region {the left-most line on the plot that rises at a 45° angle -- where $\Delta \log P(k) \sim \Delta \log k$}.

Then, at higher wavenumbers, k, the CMB P(k) rolls over and agrees with "2dF" (the "Two-degree-Field Galaxy Redshift Survey") and with galaxy clustering. The peak at roll-over is at $\lambda_{max} = 350/$ (Mpc/h, $h \sim 0.7$) and is related to the epoch when radiation density dominance gave way to matter density dominance.

What is the wave-length λ at the peak? 1 Mpc/h ~ 5 M ℓy. So the value 350 means 5 Mℓy/350 ~ 14,000 light years wave-length. Notice that the top lambda scale decreases to the right while the lower k scale gets bigger to the right (as it should since $k = 2\pi/\lambda$).

The **Cycloid equations** (exercise 4.1 p 123) describe a closed spherical Friedman matter-dominated or radiation-dominated universe from big bang expansion to a collapsing **final crunch** in terms of a time arc-angle parameter (which is usually called eta η – if ϕ and θ are space angles, why not one also for time?). Our book also uses these cycloid equations to describe the development of more local regions of "over-density" inside the universe.

Redshift-space Distortions (RSD's, p 125, see Wikipedia) "are an effect in observational cosmology where the spatial distribution of galaxies appears squashed and distorted when their positions are plotted as a function of their redshift {z} rather than as a function of their distance" (and there is a reference refers to our book, Serjeant). It is due to peculiar velocities outside of the usual Hubble flow. "RSDs have to be considered in any analysis that uses galaxy redshifts to make cosmological measurements."

Exercise 4.3 Magnitude apparent brightness: In modern terms, this 2000 year old Greek system uses a magnitude range of five to stand for 100 x in intensity (watts/m^2). So, one magnitude step is $100^{1/5}$ change in brightness ~ i.e., times 2.5 . We therefore express a difference of two magnitudes as $m_1 - m_2 = -2.5 \log_{10}(S_1/S_2)$ where more negative means brighter. The sun has apparent magnitude m = -27, Sirius is -1.46 (the brightest star), Vesta has m = +5. Filters can be used for transmitted color so that the filter U for ultraviolet centers at 364 nm, Blue B is 442 nm and V for Visual is 540 nm.

"Apparent magnitude" is usually understood, but there is also an "absolute magnitude M, of a star or astronomical object that is defined as the apparent magnitude it would have as seen from a distance of 10 parsecs (about 32.6 light-years). The absolute magnitude of the Sun is 4.83 in the V band (green) and 5.48 in the B band (blue).

In **figure 4.6** on page 130 (and more on page 154) focus on the top left-side of the graph where young hot galaxies emit most of their light in the ultraviolet! Notice that this bias is gone in older galaxies.

Balmer Series (1885) p. 132: There are four popular hydrogen photon emission series. If an orbiting electron from some hydrogen principle quantum number decays down to the lowest n = 1 state, we say that we have a "**Lyman**" series (e.g., Lyman alpha has n=2 → n = 1 {or orbital 2p→ 1s} emitting a UV photon with λ = 121 nm. The Lyman series is way too UV for our vision; but, after cosmic redshifting over a long distance, they can be visible (see pg 253). The more immediately visible Balmer series drops an electron from an excited state down to n = 2 as a lowest chosen level: so n = 3→ n=2 is called Balmer Hα with energy 1.89 eV or λ =656 nm (red color spectral line). A bigger drop from n = 4 → n = 2 is Hβ at 2.55 eV or 486 nm (blue). For higher numbers like n = 9→2 at λ = 383, the spectral lines become very closely spaced – a continuum called the **Balmer jump** or Balmer break (pg. 133 and Fig. 4.6).

The same thing happens with high n for the Lyman series too and is called the Lyman Jump.

Then there are the Paschen series down to n = 3 (infrared) and the Bracket series down to n = 4 as a selected lowest level. Being close to hot stars can ionize inner orbital electrons away from atoms thus creating orbital vacancies for subsequent series decays. Galactic dust absorbs the blue Hβ lines more strongly than the red Hα lines, and that provides a handle for deducing levels of dust. Exercise 4.4 is a long calculation dealing with assumptions versus estimates of dust attenuation.

Optical depth measures the attenuation of the transmitted radiant power through something, $\tau \propto A_v$ ["V" meaning visible (like green), *not Violet*]. Hα photons have a lower optical depth than Hβ or UV photons—red transmits better than blue. Flipping the wavelengths 1/λ {in microns} is near the number 2 on the x-axis of **Fig. 4.8** with more UV (and more attenuation) progressing to the right.

Luminosity pg 135: The Schechter luminosity function provides a parametric description of the space density of galaxies as a function of their luminosity. (p. 146 also mentions ϕ_*). https://en.wikipedia.org/wiki/Luminosity_function_(astronomy)

Page 137 **g-band** ? : **There are many color filter conventions**, and our book seems to assume that we might already know them. In today's astronomy, we often now refer to

"**SDSS:" the "Sloan Digital Sky Survey"** – as a major reference (80 million catalogued stars and galaxies). And we sometimes wish to go beyond the colors U, B, V.

"SDSS measures magnitudes in five different colors by taking images through five color filters. A filter is a kind of screen that blocks out all light except for light with a specific color. The SDSS telescope's filters are green (g), red (r), and three colors that correspond to light not visible to the human eye: ultraviolet (u), and two infrared wavelengths (i and z). On SkyServer, the five magnitudes (through the five filters) of a star are symbolized by u, g, r, i, and z. The astronomers who planned the SDSS chose these filters to view a wide range of colors, while focusing on the colors of interesting celestial objects."
https://skyserver.sdss.org/dr1/en/proj/advanced/color/definition.asp
So, g = green. Figure 4.17 page 143 uses "i-band" infra-red.

From 1998 to 2009, "SDSS used a dedicated 2.5 meter wide-angle optical telescope and observed in both imaging and spectroscopic modes. The imaging camera was retired in late 2009, since then the telescope has observed entirely in spectroscopic mode."

A list of conventional "bands" is found at;
https://en.wikipedia.org/wiki/Apparent_magnitude

Astronomical Ionized Spectral Lines: pg 141 mentions **OIII** which is doubly ionized oxygen. Singly ionized oxygen is O II and singly ionized nitrogen is N II. There is also an O I at 630nm (not ionized uses Roman numeral I). The classification of stellar spectra uses the temperature hierarchy **O B A F G K M** (Annie Jump Cannon at Harvard) – "from the hottest blue O stars to the coolest red M stars." "The visible spectral lines of singly ionized calcium (Ca II) are most intense for K0 stars (T_e = 5250 K)" {from my Astrophysics textbook}.

[And even more special color filters] **Page 146 BzK galaxy**: "A set of broad-band and narrow-band **infrared filters** was required for use with the 8.2-m Subaru Telescope and the 8.0-m Gemini North Telescope" (2001). "**BVRIzJHK imaging** with B, z, R, I, Js , Ks , K filters." "In infrared astronomy, the K band is an atmospheric transmission window centered on 2.2 μm (in the near-infrared 136 THz range)." The center for the J filter is 1.25 microns. Broadband Z is 1.033 microns. "J-band" and "H-band" are mentioned in Figure 4.23. In astrophysics, a BzK galaxy is a galaxy that has been selected as star-forming or passive based on its photometry in the B, z, and K photometric bands [WIK]
"The AB magnitude system is an astronomical magnitude system. Unlike many other magnitude systems, it is based on flux measurements that are calibrated in absolute units, namely spectral flux densities." https://en.wikipedia.org/wiki/AB_magnitude

Page 149 (and p 154) refers to the Fundamental Plane (from 1987. First glance back to page 98). Out of 4 measureable variables: Luminosity L, effective radius r_e, mean surface brightness $\mu = \langle I_e \rangle$, and velocity variation σ_{vel} , only three are independent. So, for example, $r_e \propto \sigma^{1.34}/\mu^{0.82}$, or $L \propto \sigma^{3.5}/\mu^{0.7}$.

Page 151: There was a time when astronomers did not know about huge superclusters and big voids in the fractal structuring of the cosmic web. **Pencil beam** analysis of redshifts in a sub-degree-squared area of the sky revealed a big void in 1981 and later studies: the Bootes SuperVoid (330 million light years diameter).

Page 154: The **Hertzprung-Russel** Color-Magnitude diagram is usually drawn as Luminosity Magnitude on the y-axis and spectral type or color on the x-axis. It shows main sequence lifetimes increasing with cooling down to the right.

Originally, the x-axis was the sequence OBAFGKM with hot "O" blue left and cool "M" red on the right. That means that the x-axis could also be laid out as temperature decreasing to the right. Still other plots use color index (B-V) from B-V=0 to about B-V = +1.5 or so. {Why positive? – because the magnitude scale appears to work 'in reverse', with objects with a negative magnitude being brighter than those with a positive magnitude}. See: https://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram

Figure **4.30** reverses the axes: the up y-axis is color "U – V" (instead of $B\text{-}V_{visual}$) and the x-axis is magnitude with the main interest now being population shifts with **redshift** z.

In history, spectral types began logically with a "type A" having the strongest broadest hydrogen lines. This was followed by "type B" with weaker lines. And then somehow "O" was the weakest. But then it was discovered that this weakness was due to ionization of hydrogen due to high temperatures, and a more natural order was by temperature from hot to cold. So, temperature O > B > A .

Note: A reviewer of Serjeant's book called it "a graduate school level presentation of the topic… geared to giving research level descriptions of the topics researchers in those areas would understand." This was not Serjeant's stated intention. He says it is "fully self-contained" for (what we call undergraduate seniors) interested in future PhD study but that students should look at his suggested reading sources. He assumes some previous general background in astronomy and astrophysics without which supplemental outside readings would then be necessary.

Book for Discussion: **An Introduction to Modern Cosmology**, Andrew Liddle, Wiley

At the beginning of each monthly book group at the Boulder Library, we examined some key summary concepts prior to open discussion (shown here). And before meetings, Bill Daniel provided written solutions to our homework problems and special "Notes on An Introduction to Modern Cosmology."

November, 2015 #1. The spatial metric forms for general relativity (GR) initially look strange until one examines some for simple cases (like on the surface of a basketball, sphere, $\mathbf{S}^2$).
#2 We are often told that creation from nothing preserves nothing in the sense that total expanding cosmic mass-energy is balanced by negative gravitational potential energy (a new concept for most of us—not discussed in classes).
#3 The Friedmann cosmology equation came from hairy GR—but the concepts can instead be quite simple from Newton mechanics.

December: #1: one simple equation based on red shift factor can include a lot of cosmology.
#2: Distance measures like Dc = ∫ cdt/a(t) initially seem wierd– they project old distances to our current era (e.g., a = 1/3 means "multiply by 3" = 1/a).
#3 The most common measure is "Look Back Time" (or cΔt distance in ℓyr) framed by "emit distance" and comoving distance: De<Dℓt < Dc (see Whittle p 47 problem #4). A

picture shows how to project elements of distance from small early scale factors a(t) to the present a = 1.

January, 2016 #1: Comoving distance is ruler distance at a fixed time scale a(t).
#4: Liddle problem 7.5: When (after big bang, abb) did the universe begin to accelerate (Ans: ~7.5 Gyr abb when z = 2/3 and Ωmatter also ~ 2/3). Google points to convenient cosmology calculators (like Kempner.net). In terms of deceleration, matter only counts with half strength when compared to Λ .

February, 2016: #1 Galaxy rotation curves tend to have flat velocity profiles well past luminous stars due to the increasing effects of dark matter halos. Find total galactic density versus radius, ρ(r).
#2. We live at a cosmic coincidence time where Hubble time = 1.00 t (abb), and previous deceleration and acceleration mainly balance out.
#3. When was matter-radiation equality in density? (Ans: about 88 kyr).
#4. A general formula to find cosmic time t versus z or cosmic temperature T.
#5 When did matter density = dark energy Λ density? (about 10.3 Gyr).
#6. Reference equations using simple proportionality (rather than numerical integration).

March, 2016. #1 When was the Higgs symmetry breaking and the electroweak phase transition? (just apply the radiation reference equation to get t ~ 0.1 ns after birth).
#2 Somehow the early universe produced more matter than antimatter so that their mutual annihilation into gamma rays left a tiny portion of residual matter, $\eta \sim 6\text{x}10^{-10}$ ~nB/nγ. Is it possible to produce this inside the standard model in the electroweak era? What is a "sphaleron"? We may have to go beyond the standard model for answers.

Notes for "Interstellar" for April, 2016, DP.

A standard **embedding** diagram is shown below for a massive star just shy of being a black hole (Ludwig Flamm 1916). A 5th unreal artificial dimension is added just to show curvature [vertical lift z = z(r) embedding formula {MTW}]. Rubber sheet deformation is often shown for Newtonian gravity. But it is **not** due to space but rather to curvature of time! dt/d(tau). Let time bend just a little: metric $d\tau^2 = (1\text{-}2GM/c^2r)\, dt^2$. Newtonian potential ϕ =-GM/r , so $(dt/d\tau)^2 = 1/(1\text{-}2GM/c^2r) \sim (1+2GM/c^2r) = (1\text{-}2\phi/c^2)$-- perceived flow of time versus proper time. In the book, Kip Thorne treats embedding as if it were real Randal/Sundrum (RS) 5th dimension of bulk off a 4-d Brane! RS is incredibly popular, but it is based on anti-de-Sitter space, AdS^5 like the astoundingly cited Maldacena paper AdS/CFT. But OUR universe is NOT AdS—it is becoming just deSitter due to Λ dominance. For details see: RS Model, 2006, https://www-thphys.physics.ox.ac.uk/people/MaximeGabella/rs.pdf.

Schwarzschild Wormhole [MTW p 837]: ("Einstein-Rosen bridge") described as a "paraboloid of revolution"—so how do we get Thorne's extended long throat? Also see Kip's Black Holes and Time Warps book, 1994. The latest reference on wormholes is: "From the **Flamm-Einstein-Rosen bridge** to the **modern renaissance of traversable wormholes,"** Francisco S. N. Lobo, http://arxiv.org/pdf/1604.02082.pdf. Prior to that was: M. S. Morris and K. S. Thorne, "Wormholes in spacetime and their use for interstellar travel: A tool for teaching General Relativity," Am. J. Phys. 56, 395 (1988). And CU's Homer Ellis "traversable wormholes" = "drainholes" (1973). A modification of std GRT is required!

Electron Degeneracy Pressure and White Dwarf Star Collapse

Basic Concepts in Physics page **290** has a **really NICE** "sketch" of the transition from White Dwarfs (WD) to Neutron Stars in units of sun-masses. But missing steps make hard reading: Begin with Uncertainty $\Delta p \sim \hbar / \Delta x$. Electrons are fermions so density n_e=(1 e^-) in its own little cube of side Δx and volume $(\Delta x)^3 \sim (\hbar/p)^3$. So $p = \hbar\, n^{1/3}$. Max speed of the outer electrons in WD is ~c (relativistic). So E_{fermi} ~pc [eqn 9.26] (but low mass WD's still use $E_f = p_f^{\ 2}/2m_e$). WD Mass is ~ the sum of its baryons, Nm_{Baryon} ~ Nx1.7x10^{-24} grams-- note CGS units in astrophysics. Gravitational energy is **negative**: (calculus exercise) $E_{ball} = -3M^2G/5R \sim -GM^2/R$. $M_{max} = Nm_B \sim 2\times10^{57}\times1.7\times10^{-24} \sim 10^{33}$ g. SUN $M_{\odot} = 2\times10^{33}$ g, so M_{max} ~ 1.7 suns (book says 1.85 > 1.44 for Chandrasekhar limit). At this limit, we progress to neutron stars with neutron degeneracy pressure.

Simplest example of identical particle statistics.

Repeatedly throw two coins to get combination states of HH, TT, HT, and TH with probability ¼ each. **If** these coins were bosons, then we would only count HH, TT, and {HT} with statistical weight of 1/3 for each. With Fermi-Dirac (FD) "fermion" coins we only get {HT} =100%, and the weights of HH and TT = 0 by exclusion principle.
That is, Bose-Einstein (BE) allowable states must be symmetric under particle exchange: so {HT} = (HT+TH)/√ 2. And with FD, we have **anti**-symmetry under exchange with just one state: (HT – TH)/√ 2 [e.g., heads would have (HH - HH)=0].

Cosmology Questions October 2019

Dave Peterson, 10/3/19 -10/9/19

Noether's Theorem

(dp, 5/23/08):

Noether's theorem from 1915 states that "To every differentiable symmetry generated by local actions, there corresponds a conserved current." If there is a symmetry of a Lagrangian under changes in a variable s so that dL/ds = 0, then there exists some property of the system "C" such that dC/dt = 0.

"Emmy Noether's theorem is a profound reinterpretation of the Euler-Lagrange equations." It has also "been elevated to one of the first principles of physics." "We nowadays believe that all the conservation laws come from continuous symmetries of the fundamental interactions, and so symmetry studies are essential to the understanding of all physical forces [Dick]." "The "Action", S, is an integral of a Lagrangian over fixed end points which can be expressed in generalized coordinates, q {for example, q = x}:

$$Action\ S = \int L(q,\dot{q},t)\ \ where\ \dot{q} = \frac{dq}{dt}\ \&\ p = \frac{\partial L}{\partial \dot{q}}, \qquad then\ \delta S \sim 0 \Rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = \frac{\partial L}{\partial q}.$$

That is, Hamilton's principle of stationary action δS ~ 0 implies the Euler-Lagrange equations. Consider the simplest case L = K-V = ½ mv^2 – V for a classical particle. If the potential energy V(q) = 0, then there is no q or x dependence, and L is invariant under translation in space. Then dL/dv = mv = momentum = p; so dp/dt = 0, and momentum is conserved.

The Euler-Lagrange equation for a point particle simply says that dL/dq = F = dp/dt = ma -- Newton's second law.

Or, if L = K –V doesn't depend on time, then dL/dt = 0 = dK/dt says says that kinetic energy is conserved.

A paper called "Noether's Theorem in a Nutshell" [JohnB] says, "suppose the Lagrangian L has a symmetry, meaning that it doesn't change when you apply some one-parameter family of transformations, s, sending q to some new position q(s)"

$$\frac{dL(q(s),\dot{q}(s))}{ds} = 0.\ Let\ C = p\frac{dq(s)}{ds},$$

$$\frac{dC}{dt} = \dot{p}\frac{dq(s)}{ds} + p\frac{d\dot{q}(s)}{ds} = \frac{dL}{dq}\frac{dq(s)}{ds} + \frac{dL}{d\dot{q}}\frac{d\dot{q}(s)}{ds} = \frac{dL}{ds} = 0$$

using the chain rule. So, dC/dt = 0 means that C is a conserved quantity.

Then, if the parameter, s, is chosen to be s = q, C = pdq/dq= p is conserved. Or, if parameter s = time = t, then C = pdq/dt = mvv = mv^2 is conserved – that is, Kinetic Energy is then conserved.

And, physics is the same for all angles of rotation implies conservation of angular momentum.

Noether's theorem has had applications in classical physics and field theory, but it had its origins in the apparent lack of energy conservation in general relativity theory, GRT. GRT might now be called a gauge theory whose Lie symmetry group is the group of all continuous transformations with continuous derivatives (general coordinate transformations). Emmy Noether "demystified physics" by clarifying the understanding of the principle of energy conservation as due to a symmetry under time translations. In GRT there is no "principle of local energy conservation for regions of spacetime in which there exist gravitational fields."

In QFT, Noether's theorem becomes the "Ward-Takashashi" identities. For example, phase invariance implies conservation of electric charge {the U(1) Lie group in E&M implies a divergence free current and from this one proves charge conservation."

A critic says, "The conservation laws are not explained by the symmetries anymore than the symmetries are explained by the conservation laws. In the end, Noether's Theorem is a tautology" [Mathis]. Conservation laws originally came from Newtonian mechanics which also motivated Lagrangian and Hamiltonian mechanics.

For the case of a free particle, S = ∫ Kdt = KΔt, and energy is conserved.

$$K = \frac{mv^2}{2} = \frac{pv}{2}, \qquad S = \frac{pv\Delta t}{2}\ , but\ v\Delta t = \Delta x, so\ S = \frac{p\Delta x}{2}$$

Δx = constant, and stationary action, S, means p is conserved. Also consider a particle in orbit with no applied torque, no acting force. "Rotational symmetry of space is related to the conservation of angular momentum (also symbol "L") as an example of Noether's theorem. Here

$$L = rp = rmv = \frac{mr^2v}{r} = I\omega.\ K = \frac{mv^2}{2} = \frac{I\omega^2}{2} = \frac{L\omega}{2}.\ S = \frac{L\omega\Delta t}{2} = \frac{L\Delta\theta}{2}$$

But modern physics goes well beyond Newtonian mechanics, and Noether's theorem and action principles still apply there. The action concept seems to be more general and basic. It may also have a better foundation from quantum mechanics in the sense that counting Planck pulses may be fundamental. The Feynman "Path Integral" quantum mechanics may be basic.

Cosmo Group Topic (10/3/19) : Failure of energy conservation in general relativity cosmology and Noether's Theorem:

For a rich history, see Noether's Discovery [Byer in References].

Note that Emmy Noether was a pure mathematician in the field of abstract algebra with little concern for particular applications; and she worked on this problem as a special request from David Hilbert. She only became popular more than 40 years later due to Yang-Mills/Gauge theories, but her theorems are now considered to be a key to modern physics.

The following is in response to a question by Bill Daniel – it is pretty hairy stuff beyond our usual discussions:

Bill defines Noether's theorem as: "If a continuous symmetry transformation, ϕ , only changes the Lagrangian by the addition of a 4-divergence, then there exists a Noether current that is conserved if ϕ obeys the equation of motion. The conserved "Noether charge" is carried by the conserved Noether current up to a divergentless vector field. In special relativity (SR), an example of a conserved Noether current is the energy-momentum tensor, $T^{\mu\nu}$, and its divergence is $T^{\mu\nu}{}_{,\nu} = \partial_\nu T^{\mu\nu} = 0$. Noether's theorem (1918) focuses on **symmetries of the Lagrangian** and variations of its action.

Wikipedia says, "The stress–energy tensor (of special relativity) is the conserved Noether current associated with spacetime translations. The divergence of the non-gravitational stress–energy is zero. In other words, non-gravitational energy and momentum are conserved." $T^{\mu\nu}{}_{,\nu} = 0$ (just comma for ordinary partial derivative needed for SR). Energy conservation goes with "symmetry under time translations."

Notation and definitions: "Divergent-less" = divergence-free = $\nabla \cdot F = 0$ or $\partial_i F^i = (\partial / \partial x^i) F^i = 0$ for the case of vector fields F (i = 1,2,3). An example is the magnetic field $B = \nabla \times A$ where A is the vector potential, and div curl = 0. So, B is divergent-free.

By Gauss' Theorem (or divergence theorem),

[$\int_V \nabla \cdot F dV = \int_S F \cdot n dS$*, Surface* $S = \partial V$*].* The flux integral of a divergenceless vector field over a closed surface is $\int F \cdot n dS = 0$; or the flux through a closed surface is zero. For a "4-divergence" we say $\partial_\mu F^\mu = 0$ (μ = 0,1,2,3). Beware that in general relativity, divergence should mean covariant divergence which can differ from ordinary divergence because of curvature factor additions (Christoffel symbols, Γ). This can be a confusion factor.

The energy-momentum tensor **T** = $T_{\mu\nu}$ is a 4d generalization of an older Newtonian 3d stress tensor "used for stress analysis of material bodies experiencing small deformations." The time component of the energy-momentum tensor is the energy density, T_{oo}, and the T_{oi} terms are momentum densities. We might care about some pressure terms T_{ii} but not often about stress itself. Being "conserved" implies a zero time derivative, $dT_{oo}/dt = 0$. In relativity, we often care not just about energy but the whole energy-momentum 4-vector (E, **p**).

Combining the last two paragraphs: Conservation of 4-momentum can be expressed using an integral formulation as well as the differential formulation, $\nabla \cdot T$ (using commas , in SR). A space-time volume about an event has a boundary surface ∂ Vol . "Every bit of 4-momentum which flows into V through ∂V must somewhere flow back out" – no interior sinks. [MTW p. 143]

Noether Symmetry in GR here refers to general relativity being a gauge theory with Lie symmetry group of all continuous coordinate transformations. Symmetry refers to the symmetry of Lagrangians, L = L(t, q, dq/dt) where q is a generalized coordinate (could be x,y,z). If there is no

dependence on time in L = L(no t), then energy is conserved: $dE/dt = dH/dt = dL/dt = 0$ where H is the Hamiltonian ($H = \Sigma p_i(dq/dt)^i - L$).
If there is no dependence on space coordinates, q, then momentum is conserved ($p = \partial L/\partial (dq/dt)$). The continuous coordinate transformations can just be : $x \rightarrow x+\delta x$ and $t \rightarrow t+\delta t$. Noether's work applies to Lagrangians with an action from Hamilton's principle: action $S = \int L\, dx^n$ being an extremum. Solving $\delta S = 0$ yields more useful **"Euler-Lagrange equations** that are easier to interpret**.** In the case of Newtonian mechanics, The Lagrangian is simply L = KE – V, and the Euler-Lagrange equations result in Newton's law, F = ma.

For general relativity, the matter Lagrangian is $L_m = -mc^2\, d\tau/dt$ where tau is proper time. But, when expanded, its variation is $\delta L = (½)m\, \delta[g_{\mu\nu}\, (dx^\mu/dt)(dx^\nu/dt)]$.

The total Lagrangian is $L = L_m + L_G$ where L_G expresses <u>G</u>eometry (gravitational curvature), $L_G = R\sqrt{-g}\,/16\pi$ (for c = 1, G_{Newton} = 1 units).

Question: what is the divergence of **T** in general relativity (GR) and its meaning for the conservation of energy.

The Einstein field equations (EFE) in condensed form are **G** = 8π**T** (with Newton's gravitational constant G = c =1, $G_{\mu\nu} = R_{\mu\nu} - ½\, Rg_{\mu\nu}$ and **T** = $T_{\mu\nu}$). Geometry (on the left side) tells matter how to move, and mass/energy tells geometry how to curve. A primary goal of the equation is to calculate the space-time metric $g_{\mu\nu}$. Both G and T are divergence-less (covariant divergence free).

The **<u>Einstein-Hilbert action</u>** is $S = \int L_G\, d^4x = (1/2\kappa) \int R\sqrt{-g}\, d^4x$ where $\kappa = 8\pi G/c^4$, <u>R = the Ricci scalar</u>, "-g" is the |metric determinant| , and for real problems we desire the variation of action to be zero, δS=0. Note that Einstein kept on missing the 2nd term of **G** until achieving the final form in 1915 (concurrently with David Hilbert). In our expanding universe case, $\sqrt{-g}$ may be ~ positive $a^3(t)$.

For $S = \int d^4x\sqrt{-g}\, (L_m + R/\kappa)$, δS=0 results in the field equations **G** = 8π**T** or **T**/κ (this is how David Hilbert got his field equations, really $\delta S/\delta g_{\mu\nu}$ – variation wrt the metric tensor). For $S = \int L_m\, d^4x$, δS=0 implies the energy momentum tensor, **T**.

The central problem for GR is that divergence has to be covariant divergence: $T^{\mu\nu}{}_{;\nu} = \nabla_\nu T^{\mu\nu}$ (with a semicolon for derivative). That adds Christoffel symbols to the divergence in the SR non-curved-spacetime case: $T^{\mu\nu}{}_{;\nu} = \Gamma^\mu{}_{\sigma\nu} T^{\sigma\nu} + (\sqrt{-g}\, T^{\mu\nu})_{,\nu} / \sqrt{-g}$.
That has implications for the flow of momentum-energy through volumes. In the curved space of general relativity, we have Levi-Civita parallel transport depending on transportation path so that <u>"flux is not well defined."</u> Now, we can always find a special coordinate frame reference so that Γ =0 in a locally Minkowski space; and then the flux is defined (like for a person in free fall not feeling gravity). But that only works locally and not in the coordinate independent fashion intended in GR. "Reality" is almost defined as "invariant." We have to have a selected fixed coordinate system to get an energy conservation law in integral form [Baez]. Another way of saying that is "in the FLRW models, there is no time-like Killing vector and thus no kind of conserved Killing energy or matter content."

To express the divergence of **T** in GR using only partial derivatives (commas), one introduces a "stress energy pseudotensor" $t^{\mu\nu}$ such that an effective $T^{\mu\nu}_{eff} = T^{\mu\nu} + t^{\mu\nu}$. The divergence using t is now equivalent to the covariant divergence:
$(T^{\mu\nu}_{eff} = T^{\mu\nu} + t^{\mu\nu})_{,\nu} = 0$ is equivalent to $T^{\mu\nu}{}_{;\nu} = 0$. Then we can convert back and forth between volume integrals and surface integrals [MTW p 465].

By bypassing the covalent divergence, none of the above terms involving t have any "geometric coordinate-free significance." "Because $t^{\mu\nu}$ are not tensor components, they can vanish at a point in one coordinate system but not in another." A localized energy density t^{oo} for the gravitational field is ambiguous. However, certain volume integrations can be meaningful in "asymptotically flat regions far outside the source." "The mass-energy of a neutron star is less than the mass-energy of the same number of baryons at infinite separation." "No Γ's means no gravitational field, and no local gravitational field means no local gravitational energy-momentum."

Carroll says div T=0 applies when "spacetime is standing completely still" but not when it is evolving (as in the expanding universe). Time symmetry is broken, and Noether's theorem should not apply to our changing universe. Landau said that by ignoring the gravitational field in T, div T = 0 "does not generally express any conservation law whatever." Tamara Davis [Sci.Am,July,2010] says that in the expanding rubber balloon metaphor "we are free to consider this relative motion as expansion of space OR movement through space." Actual movement and Doppler shifting can allow one to reclaim energy conservation (the majority doesn't believe this).

Alternatively, a cosmology based on the Hamiltonian would demonstrate energy conservation throughout, but this is not generally accepted. One such model [Ibison] treats the scale factor a(t) as a dynamical variable, treats Hubble flow like da/dt as kinetic energy, treats (includes) the gravitational field as a negative energy density, and consistently demonstrates energy balance in a given coordinate system. Although probably wrong, the paper does offer many stimulating thoughts.

References for Noether:

[Byer] Nina Byer, "E. Noether's Discovery of the Deep Connection Between Symmetries and Conservation Laws," 1996.
http://cwp.library.ucla.edu/articles/noether.asg/noether.html or
http://www.physics.ucla.edu/~cwp/articles/noether.asg/noether.html

[MTW] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler, GRAVITATION, Freeman, 1973 {1279 pages}.

[Ibison] M. Ibison, "Hamiltonian Cosmology," https://arxiv.org/abs/0807.1884 (not accepted for publication in Classical and Quantum Gravity).

[WIK] Wikepedia is a wonderful source of information on topics. E.g., ttps://en.wikipedia.org/wiki/Noether%27s_theorem

[Baez] http://www.desy.de/user/projects/Physics/Relativity/GR/energy_gr.html Is Energy Conserved in General Relativity? [And, yes, John Baez is related to Joan Baez, his uncle is Joan's father, physicist Albert Baez].

[JohnB] John Baez (2002) http://math.ucr.edu/home/baez/noether.html

[Dick] Auguste Dick, http://www.cscs.umich.edu/~crshalizi/reviews/dick-on-noether/

[Mathis] Miles Mathis http://milesmathis.com/noeth.html

This note was in response to my previous response to a question from the Boulder Cosmology group:

Gravitational Energy (?)

Is energy lost in an expanding universe during cosmological red shifting? Some say yes and some say no. One has to define what is meant by "energy" and by "conserved." The total energy of the universe may be undefinable, and conservation of energy may lie outside of the laws of general relativity.

A first conceptual problem is that energy is seen differently in different moving frames of reference. For example, a 2 kg ball moving at 10 m/s to the right has a kinetic energy of ½ mv2 = 100 joules – and one can see the KE by having the ball hit and indent a soft clay target. But in a different frame of reference of a lab moving to the right at 5 m/s, the KE seen is now seen as just 25 joules. And the ball hitting the moving clay only partially indents it. Here energy is not lost, it is just seen differently.

This is also true for light even though it always has speed c. There is Doppler shifting seen in a moving frame of reference – but again energy is not lost but just seen differently in the two frames. In an expanding universe with different cosmic flows, there are an infinite number of different inertial frames with different speeds and directions. It is challenging to keep track of all of them and harder still to think of summing them all up. In General Relativity we have a new problem with accelerating frames. We feel a gravitational field. But if we are in free-fall accelerating downwards, the field goes away. One cannot state or localize the energy of a gravitational field --{an “energy psudo-tensor” is not a coordinate invariant tensor}. And red-shift from expanding space is not the same as Doppler shifting due to relative motion—“the expansion of the universe does not consist of objects actually moving away from each other - rather, the space between these objects stretches.”

One article says: “Does General Relativity offer a possible violation of energy conservation? The scary answer is maybe, actually. There are a lot of quantities that General Relativity does an excellent and precise job of defining, and energy is not one of them. In other words, there is no mandate that energy must be conserved from Einstein's equations; energy is not defined by General Relativity at all!

… as the Universe expands, photons lose energy. But that doesn't mean energy isn't conserved; it means that the energy goes into the Universe's expansion itself, in the form of work { I don’t like this sentence, where are the pressure and forces pushing against “what?”—does space have inertia?}. Steve also expressed worries about this idea.

For reference see: https://www.forbes.com/sites/startswithabang/2015/12/19/ask-ethan-when-a-photon-gets-redshifted-where-does-the-energy-go/#1984db734891

Also, one can also do a thought experiment where a [massive closed] universe contracts after expanding and reverses all that “lost” energy back to the original energy (or energy density).

Another article elaborates: “It turns out that in Einstein’s theory of general relativity, regions of space with positive energy actually push space outward. As space expands, it releases stored up gravitational potential energy, which converts into the intrinsic energy that fills the newly created volume. So even the expansion of the universe is controlled by the law of energy conservation. {these views are interpretations-- and they seem to vary }.

But then there is Sean Carroll: “When the space through which particles move is changing, the total energy of those particles is not conserved” (in part because vacuum energy is a constant and we are accumulating more and more of it).
http://www.preposterousuniverse.com/blog/2010/02/22/energy-is-not-conserved/

{versus Scientific American,
https://people.smp.uq.edu.au/TamaraDavis/papers/SciAm_Energy.pdf}.

And previous:

https://en.wikipedia.org/wiki/Chronology_of_the_universe#Electroweak_epoch
that tells what particles are present when. I believe the inflation period is all fields (no familiar particles).

Electroweak symmetry breaking (EWSB) occurred when the Higgs field developed a vacuum expectation value (VEV ~ 246 GeV, 10^{-12} sec ABB, “cosmic time” after the big bang). It is at this pico-second time that photons and quarks can be said to exist. Within the confines of the standard model, this symmetry breaking <u>cannot</u> explain baryogenesis asymmetry (EWBG) – some extensions or physics have to be added to the SM (there are many suggestions).

The quark-gluon plasma transition is near 175 MeV (above that kT- energy is plasma, QGP at $z\sim10^{12}$, t~10μs). Below this energy, Hadrons now begin to exist. At 10 seconds, anti-matter is gone and black body photons dominate the universe. Big Bang nucleosynthesis occurs at 10s-20 minutes ABB. The deuteron binding energy is 2.2 MeV, so scale must be near $z\sim 10^7$. The first black holes were about 300 Myr ABB (z ~ 13) after the creation of giant stars.

String theorists say they have verified Hawking entropy for black holes (-- I'm a skeptic).

Friedmann Equations

Dave, 9/6/19- 9/8/19 [Re: ongoing discussion about the equations].

We refer to two Friedman equations, but there are three equations that go by that name (along with a "fluid equation" expressing conservation of energy). Only two of the three equations are independent, and the choice of stating which two varies from book to book.

The name "The Friedman Equation" (singular) is:

$$(dR/dt)^2 = 8\pi G(\rho_m+\rho_r)R^2/3 - kc^2 + \Lambda c^2R^2/3$$

[e.g. Serjeant text eqn.1.7, the "first" equation]. When divided by R^2, the term on the left becomes $[dR/Rdt]^2 = [da/adt]^2$ = Hubble's $H^2(t)$. I'll call this equation F_1. In all books and articles , it is accompanied by a second independent differential equation also called a Friedman equation that can appear in two forms: F_{acc} or F_{dyn}, the "acceleration equation" [Serjeant eqn 1.8] or the "dynamic equation" below. The acceleration equation for d^2R/dt^2 or R'' is more commonly shown now; but earlier sources and Friedmann himself used the dynamic equation shown here [e.g., MTW, Nuss].

DYN: $\mathbf{2R''/R + R'^2/R^2 + kc^2/R^2} = \lambda$ (and/or) $-8\pi GP/c^2$, **is** eqn F_{dyn}= "F_1"$/R^2$ + 2"F_{acc}"/R.

[where the prime on R' means dR/dt , $\lambda = \Lambda c^2$ and P = pressure (which is zero for a Friedmann expanding dust universe)]. One could invert this formula and solve for equation F_{acc} in terms of F_1 and F_{dyn}. The name IVE or "initial value equation" refers to how a(t) or H(t) would vary from current reference "$_o$" values of densities, ρ_{ro} and ρ_{mo} with a_o=1 using equation F_1 – an "applied" F_1. Note that F_1 can be derived by integration from F_{acc} combined with the fluid equation $d\rho/dt = -3(dR/Rdt)(\rho+ P/c^2)$. Or, F_{acc} from differentiation of F_1 combined with the fluid equation [as in exercise 1.3 p 291]. They are related by enforcement of conservation of energy.

Equivalently to linear combinations, one could solve both F_1 and F_{acc} for the term $(8\pi G\rho R^2/3)$, equate the results and get the dynamic Friedman equation, F_{dyn}, shown above. The density term is absent from this equation, but of course that doesn't mean it is zero – it can vary from about $\Omega_m+\Omega_r \sim 0.3$ (now) up to almost critical density, $\rho = \rho_c$ and $\Omega_m+\Omega_r \simeq 1$ in the early universe. It still has to be accompanied by F_1 which does contain density and curvature and Λ.

In the same way, a recent proposal (8/29/19 http://sackett.net/SerjeantNote4Page19.pdf) solved F_1 and F_{acc} for Λ, equated them and got a new equation without lambda – but this should **not** mean that $\Lambda = 0$! And the Friedman equation F_1 should still also accompany it. We started with two independent Friedman equations -- combining them together, like $F_{proposed} = RF_{acc} - F_1$, doesn't stand alone taking the place of the original two.

Bill Daniel nailed the primary problem with this proposal: you cannot have k = 0 and $\Lambda = 0$ together with $\rho < \rho_{critical}$ [Daniel]. The proposal claimed that curvature k is not determined by the Friedman equations (so one could pick k=0). Actually, that is not quite true. One might input k at will, but the resulting solutions have to be **consistent** with that choice – and that limits input freedom. The "deceleration parameter", $q \equiv -(RR''/R'^2)$ [page 23] is still often useful in determining k. {But for our universe, we anticipated something like $q \sim +½$; But, it turned out to be $q \simeq$ negative 0.55 ‼ -- Accelerating rather than decelerating ‼ Upsetting… but still flat, k=0 and having critical density. }

The relationship to k for the case $\Lambda = 0$ is:
Closed Universe (positive curvature) $k > 0$ if $q > ½$ or $\rho > \rho_c$.
Flat Space (zero curvature) $k = 0$ if $q = ½$ or $\rho = \rho_c$.
Open Universe (negative curvature) $k < 0$ if $q < ½$ or $\rho < \rho_c$.

For $\Lambda \neq 0$, we have a closed universe when $\rho / 2\rho_c > (q+1)/3$ – a little more complicated. And the universe will expand forever if $\Lambda >$ a special value $\Lambda_{critical}$.

The Friedman equations have input k, ρ, p, and Λ and output derivatives of a or R and Hubble values. q can be found from these relating to k (-, 0, +1).
$q = (1/3H^2)[4\pi G(\rho + 3p/c^2) - \Lambda c^2$, and, for now for us $q \simeq \Omega_m/2 - \Omega_\Lambda$.
[For the proposed model, $\Omega_m \sim 0.22$ and $q \simeq 0.11$ (both meaning "**open**" k<0, not zero]). The input k = 0 was not self-consistent with the results k < 0.

Why should time-curvature cause Newtonian gravitational force?:

A question from the last book club meeting was essentially, "Why should we be forced onto our chairs if Newtonian gravity is really just a curvature of time?" (April, 2016)

In Newtonian gravity, we say that force down is F = mg. But general relativity began with the principle of equivalence that one cannot distinguish locally between g down and acceleration, a, up. Also, the **proper acceleration** of an object is determined by an observer who is allowed to be in natural free fall along a "geodesic" (say there is a big hole in the ground next to the chair and someone is falling through it). He is not accelerating, but he sees the fixed chair accelerating **upwards** relative to him. "Gravity sucks," and people on Earth are accelerating against it. The fixed chair has to push up against the body on it, and that requires it to compress. We like to think of our perspective as central, but a lot of physics has been done to de-throne that view.

Although force, F, is a key concept in classical physics, it is barely mentioned in general relativity (GR). There, gravitational force is treated as a "fictitious" force that can be transformed away by free fall. But gravity feels real to us as local observers. Anyway, the equation F = ma should generally be replaced by dp/dt = ma with momentum/energy always being more fundamental throughout physics (including special and general relativity).

GR was meant to give physics from anyone's perspective (invariant with respect to position, time, velocity, and acceleration). But, practically, we do mostly care about special frames of reference. Schwarzschild coordinates, for example, are centered about a big spherical mass. In that perspective, we can talk about gravity as having a gravitational potential and being equivalent to space/time warpage about that mass, or as being replaced by an effective "index of refraction."

GR books then say that solutions of its Einstein equations reveal that gravitational acceleration, g, is proportional to the gradient of the metric time coefficient (g_oo), so that g and gravitational potential V are related to time curvature dt/d(tau) = perceived time change over local clock "proper" time change. This is valid in the Newtonian realm of slow motion ($v \ll c$) and weak gravity fields (well below that of neutron stars). Our ability to experimentally detect such time flow differences versus change in altitude, h, has gone from the 1959 Pound-Rebka gamma ray experiment of h=22 meters now down to a centimeter using atomic clocks. Can nature feel the difference down to atomic sizes? Although ordinary gravity is time curvature, massive sources do distort both time and space. And, indeed, for fast motion, $v \sim c$, particles respond to the curvature of space as well (so light gets bent equally from time and from space curvature).

An early thought in general relativity is: If we toss a stone into our air, its path will be a parabola of vertical space over time. If we change the time axis from t to ct (as in relativity), it will stretch out so drastically that the parabola becomes just an approximation to the very top of a great circle. This great circle geodesic has a huge radius of curvature of $R = c^2/g \sim$ about one light year! With a strong sideways velocity, the motion is still parabolic as an approximation to a great circle with time strongly dominating the horizontal axis. Velocity has to be really big to have any effect on the curvature: $K = 1/R$. Somehow, we (or stones) know how to go with the curvature.

Small quantum world objects like atoms and coherent molecules also fall freely in vacuum and each also possess a "basic vibration frequency" and moves so that their waves have "least action" (smallest number of waves along a path). We do not talk about forces in quantum mechanics either, but mean motions and geodesics are still relevant. The rest frequency of matter particles is huge because it is proportional to rest mass-energy, $f = (E = mc^2)/h$, which is huge. To experience gravitational potential, it seems that these objects must be able to experience time flow differences or phase alterations over tiny scales like angstroms. That stretches the imagination --but to the same degree for experiencing the old Newtonian potential (tiny differences of V over tiny scales h). But suppose each object is "exploring" the possibility of moving over distances much larger than that (a virtual exploration). That smacks of the same "teleology" of having a great multiplicity of possible Feynman paths towards a result that ends by picking just the right one.

Light has frequency and wavelength; and its path bends through glass having a refractive index, n, that can vary or "disperse" with values of nu and lambda (red goes through a prism straighter than blue). As Mike Jones notes, a gravitational field can also be thought of as having an effective varying <u>refractive index, n,</u> and that can be used to calculate bending of starlight and gravitational lensing. Increasing g acts to produce an apparent slowing of light below c like $c' = c/n$ as in Snell's law.
Then $n = c/c' \sim (1\text{-}2MG/c^2) \sim$ free-wavelength/ new-wavelength $=\lambda o/\lambda$. The idea also carries over to matter-waves as well when their "rest mass" is included. Each case has its own "dispersion relation," for energy $E = H(p,z)$, or frequency = f(wavelength) – [since $E = h\nu$ and $p = h/\lambda$]. Unfortunately, fast moving particle with mass see an index N that depends on speed. I would say that fast speed "reveals" the underlying spatial curvature. [We've seen such "velocity dependent potentials" before when dealing with magnetic fields].

It is often convenient to think of static forces in terms of compression of springs. A solid object on a solid table experiences electrostatic compression of atoms (a very high spring constant in $F = kh$). Sitting in a chair involves compression of body and muscles and maybe cushions too with a weak spring constant. In both cases, there is a vertical height displacement, h, of a center of mass opposing a change in gravitational potential. Spring force is not fictitious; it is a real thing in the realm of electrical character and structure of materials. If this resistance force were not present, one would experience no force because you would be in free fall ($a = g$)

We have several issues to contemplate: the large classical world, geodesics of motion, a "static world" of very little motion, and the small quantum world. And there is also the issue of particles with mass versus vibrating light photons without any rest mass. And all of these have to dovetail into each other for total consistency. I'm not sure that has really been done, so the little effort here may be incomplete and unsatisfactory.

With the constraint of resisting forces against motion, perhaps "stationary" small objects are also still exploring extended space but without any net effect. Then they would know which way was "down" without experiencing a "force" of gravity. Quantum objects are lacking in precise positions in space and do have some sort of virtual existence over extended localities. So single electrons or even buckyballs can seem to travel through widely spaced double slits well above an object's individual size. Unfortunately, explanations in this realm become difficult and tenuous.

2. A second question was about **Dark Matter as WIMPS** (weakly interacting massive particles): a particle interacting gravitationally with mass in the range 1 GeV to 1 TeV. It was expected to self-annihilate with a typically weak rate. Unfortunately, all the searches so far have come up empty and SUSY particles have not been seen at the LHC so that doubt now exists.
One test idea was to use **Liquid xenon** as a choice material because it is itself a scintillator: it emits characteristic ultraviolet scintillation photons when a recoiling nucleus passes through the liquid, and it is also transparent to these photons. Thus the photons act as a prompt signal of a nuclear recoil and they can also very be easily detected using photomultiplier tubes that look directly into the xenon liquid, even when there are not that many photons produced. The space of cross-section versus particle mass (GeV's of energy) is being progressively exempted.
Motivated by the neutralino, somewhere around 1990 **WIMP** idea dominated DM and millions of euros were and are still being spent in the search for WIMPs. Now, the WIMP is being disfavored, and the more favored axion is also in enough trouble that we should look elsewhere.
http://arxiv.org/pdf/1604.05207.pdf Paul Frampton.
And one place to look is in special modifications of general relativity.

3. **Comments on Basic Concepts in Physics,** from the cosmos to quarks, by Masud Chaichian, 2014,377 pages.

Our previous goal in the Cosmology book club was to obtain an intuitive understanding of modern physics primarily through reading popular books with minimal (if any) supporting math. To a large degree, we've done that. Our most challenging book was Deep Down Things that took us deeper into modern physics (but there aren't many books like that). Many of the analogies and heuristic aids we've encountered have helped our understanding but have also been over-simplified, misleading, and sometimes conceptually wrong (like gluon string as rubber band).

We've been contemplating a deeper overview with more accuracy and exposure to equations to clarify key concepts. The suggested book is a concise overview that covers a lot of territory with which we all should really be "acquainted." That is the intention, exposure to many of the basics without expectations of mastery (yet). It is for a really broad audience from freshman to post graduate, so we should not expect to derive or fully understand everything or maybe even half of it. Each of us goes as far and deep this time as they wish and key in on topics of greater personal interest (with maybe a little googling on favorite ideas). In many cases you get to "see" the relevant equations for the first time. Again, our goal is intuitive understanding more than actual ability to calculate. We will address what is important and why.

I posted a previous review at: http://www.amazon.com/gp/customer-reviews/R3GLASH60A2A7J/ref=cm_cr_arp_d_rvw_ttl?ie=UTF8&ASIN=3642195970

[And Musser's Spooky action is at: http://www.amazon.com/product-reviews/0374298513/ref=acr_search_hist_4?ie=UTF8&filterByStar=four_star&showViewpoints=0]

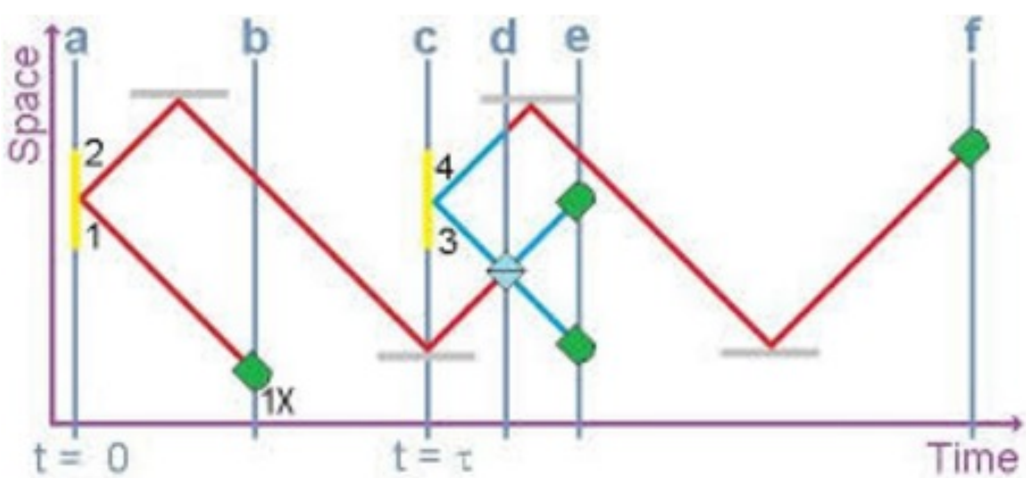
We will discuss this further before locking in the book. I've heard some reluctance (…but one must also propose alternatives…).

Thanks, Dave.

NOTES: In general relativity, gravity is a fictitious force Stack Exchange
The General Theory of Relativity was Einstein's stupendous effort to remove the restriction on Special Relativity that no accelerations (and therefore no forces) be present, so that he could apply his ideas to the gravitational force. It is a measure of the difficulty of the problem that it took even the great Einstein approximately 10 years to fully understand how to do this. Thus, the General Theory of Relativity is a new theory of gravitation proposed in place of Newtonian gravitation.
Einstein said there is no such thing as a gravitational force. Mass is not attracting mass over a distance. Instead, it's curving spacetime. If there's no force, then how do you explain acceleration due to gravity?.... Now, let's get back to those geodesics. A body undergoing geodesic motion feels no forces acting upon itself. It is just following what it feels to be a "downward slope through

spacetime" (this is how the bending affects the motion of an object). The particular geodesic an object wants to follow is dependent upon its velocity, but perhaps surprisingly, not its mass (unless it is massless, in which case its velocity is exactly the speed of light). There are no forces acting upon that body; we say this body is in freefall. Gravity is not acting as a force. (Technically, if the body is larger than a point, it can have tidal forces acting upon it, which are forces that occur because of a differential in the gravitational effect between the two ends of the body, but we'll ignore those.)…. Einstein's insight is that there is no local experiment that can tell the difference between being in a gravitational field versus being in an accelerated reference frame….Quora: So the reason why it feels like there is a force holding you to the chair you are sitting in right now is because the chair is forcing you to NOT follow the geodesic path you would otherwise have followed in the curved space-time caused by the earth. So you are really feeling the electromagnetic force of the chair's atoms pushing against the atoms of your body preventing you from traveling in a straight line - not the fictitious gravity force. If the chair were not there, you would feel no force at all - exactly as if you were floating weightlessly in space far from any gravitating object. You would feel that weightless up to the point where you happened to hit something else as you travel along your "straight" geodesic path in curved space-time.
Adler p 121: gravity as a metric phenomenon. The geodesic solution of the weak field metric gives acceleration g $\propto$ ∇ goo.

Figure: **“Experimental Entanglement Swapping: Entangling Photons That never Co-existed”** [my favorite entanglement picture]. A UV laser beam enters from the left through an SPDC crystal where it can output two lower energy entangled photons. Photon 1 goes straight to a detector while 2 goes to a Beam Splitter BS. The laser beam passes through the first crystal into a second one where two new entangled photons are created. Photon 3 and 2 coincide exactly at time =d on the BS with 2 and 3 becoming entangled and 2 then getting detected. Then 4 continues to its final detection. 1 and 4 are now entangled although 4 originated well after the creation and detection of photon 1. “TI” would say that the detection of photon 4 sends a confirmation wave back in time from photon 4 at f along a convoluted path to the sources of the offer wave thus establishing correlation. But in this case there are two sources.



Science Highlights of the Months

Dave Peterson, April, 2014 – 10/9/19

FYI: Here is a collection of leading summary notes I've recorded for each month from Physics ArXiv's, Physics World, Science News, various journals, other physics, cosmology, science sources and daily news stories (not all science).

Summaries for the Months:

Physics Note September 2019

Dave Peterson, 9/2/19 -10/1/19

Summary

1. Science 9/6/19: Radio emissions from the magnetic pole of a newly mapped pulsar, known as PSR J1906+0746 reveal time slices of changing intensity over a 20 year period (the lighthouse beam shifts direction revealing its cross section). It has a neutron star companion that makes it **precess** about 2.2 degrees per year. The beam should cease to be detectable by 2028.
2. The Terrell-Penrose effect of 1959 says that any relatively small object traveling near light speed will visually appear to be rotated but with the same apparent width—the Lorentz contraction is cancelled out. A sphere is still a sphere and a cube is a rotated cube. 60 years later, this effect is not well known.
3. What's up with **LIGO?** Wednesday, September 04, 2019 (what is needed is other independent detections like that for the N+N kilonova). http://backreaction.blogspot.com/ Sabine Hossenfelder's Blog
4. Astronomers have discovered the **most massive neutron star** to date, millisecond pulsar with white dwarf binary approximately 4,600 light-years from Earth. This record-breaking object is teetering on the edge of existence, approaching the theoretical maximum mass possible for a neutron star [J0740+6620, 2.17 times the mass of our Sun]. S. News.
5. **LOFAR** (Low frequency array radio telescope network, Netherlands, from 2012) is a massive new endeavor over 48 stations to investigate λ ~ 1.3 m to 30 meter wavelengths (10-230 MHz) www.lofar.org. It has many goals beginning with the epoch of reionization (post dark ages red shift of 21 cm line of hydrogen 1420 MHz from 6<z<10): Oct, 2018 "LOFAR Discovery of a 23.5 s Radio Pulsar," Feb 2019 "Scientists Hit Cosmic Jackpot with Discovery of 300,000 Distant Galaxies." Etc.
6. Rick Saltzman. https://www.nature.com/articles/s42005-019-0203-z "The sound of Bell states" Classical non-separability: Three parallel ½" x 2 ft Aluminum rods tied together with rubber bands and separated by epoxy or honey. 33 kHz OAM. "we can tune the eigen mode superposition, that is, the Bell state."
7. ScienceMag: Since 1970, the total number of birds in North America has dropped **by 29%—**about 3 billion birds lost in under 50 years. ‼ This includes even robins and sparrows (loss of habitat, chemicals, pesticides, loss in insects).
8. In 1968, Brandon Carter showed that the Kerr Metric was "the only stationary metric with a simply connected bounded event horizon, i.e., the only possible black hole." This wasn't known by Kerr and friends in 1963-1964. "What happens after the outer horizon forms is still a mystery after more than four decades." There is no clear interior solution, and Kerr believes that the inner event horizon never actually forms." Roy Kerr: https://arxiv.org/pdf/0706.1109.pdf

9. **Lies:** The Mexican war of 1846 began with a deliberate provocation by Polk in Texas and also the desires of slavers (outlawed in Mexico); The Spanish-American War of 1898 was due to an explosion on the USS Main that was actually an internal accidental explosion. In 1915, the RMS Lusitania blew up from internal munitions not from a German torpedo. Roosevelt enticed and … provoked Japan into WW2. The Vietnam War 1964 "Gulf of Tonkin incident" never happened. The Iraq war in 2003 was based on lies such as WMD.
10. A minor change: Most oil and gas trade associations now argue that expanding natural gas production offers the best path for reducing the United States' carbon footprint. "Sometimes getting exactly what you want is the very worst thing you … can get." "Regulation is not our enemy. It is the way we keep faith with the public."

Physics Notes August 2019

Dave Peterson, 8/1/19 – 9/2/19

Summary

1. The famous LIGO/VIRGO **neutron star merger** GW170817 GRB appears to have ended up as a 2.7 sun mass magnetar rather than a black hole. Another neutron star merger XT2 (at z= 0.74, 136 photons over 3 hours) occurred on March, 2015 (delayed analysis) and was seen only in X-rays by Chandra and also had a signature of a millisecond neutron star magnetar.
2. Schwinger disliked Feynman diagrams because he felt that they made the student focus on the particles and forget about local fields, which in his view inhibited understanding. He went so far as to ban them altogether from his class, although he understood them perfectly well
3. Physics APS PRL: Quantum interference and entanglement across an astronomical distance between the Sun and a quantum dot. With filtering of sunlight to match the frequency and polarization of the photons produced by a quantum dot, one can show the HOM effect (bosons acting together). But Bell nonlocal correlations were also demonstrated. Thermal light has an unambiguous quantum nature.
4. { https://arxiv.org/pdf/1908.09823.pdf (15 authors)} :
… introduces an Entangled Neutron Beam! where individual neutrons can be entangled in spin, trajectory and energy over about a micron, wavelength near 0.4 nm (speed~4km/s) and energies differences below a neV [Larmor RF spin flippers at ISIS pulsed neutron source in the UK]. "Spin-Path" ψ is $|\uparrow 1\rangle + |\downarrow 2\rangle$, and GHZ state is $\psi = |\uparrow 1\, E_-\rangle + |\downarrow 2\, E_+\rangle$.
5. Quantum correlated 11.1 **keV x-rays** were produced by prametric down-conversion using a very bright x-ray beamline on diamond crystal [PR X]
6. A chip made with **carbon nanotubes**, not silicon, marks a computing milestone (14,000 carbon nanotube transistors , but a long way to go yet). Carbon nanotubes are almost atomically thin and ferry electricity so well, they make better semiconductors than silicon. In principle, carbon nanotube processors could run three times faster while consuming about one-third of the energy of their silicon predecessors.

Physics Notes, July, 2019

Dave Peterson , 7/4/19 – 8/1/19

1. Beyond the middle of our galaxy at the edge of the Local Group is a **Local Void** about 200 Mℓy across. "The Local Void has a substantial dynamical effect, causing a deviant motion of the Local Group of 200–250 km s^{-1}. The combined perturbations due to repulsion from the Local Void and attraction toward the Virgo Cluster account for ~50% of the motion of the Local Group in the rest frame given by the cosmic microwave background {n~ 17,000 galaxy distances in our cosmic neighbourhood}.
2. PT March 2019: As white dwarf stars cool, oxygen nuclei settle towards the center with carbon nuclei on the outside. Electrons are a Fermi gas with degeneracy pressure opposing gravity. **The oxygen nuclei crystallize** into a body centered cubic lattice. Gaia data on 15 000 white dwarfs is on the verge of revealing latent heat of crystallization.
3. In 1971, at a Baskin-Robbins ice-cream store in Pasadena, California, Murray Gell-Mann and his student Harald Fritzsch came up with the term "**flavour**" to describe the different types of quarks. The SM is supposed to be flavor independent for all three generations of fermions.
4. **Terraforming Mars** is likely an unfulfillable dream. In 2018, a pair of NASA-funded researchers from the University of Colorado, Boulder and Northern Arizona University found that processing all the sources available on Mars would only increase atmospheric pressure to about 7 percent that of Earth - far short of what is needed to make the planet habitable. But a 3cm thick layer of aerogel could allow photosynthesis and higher temperatures on a regional basis.
5. Pre-recombination era: 63% DM, 15% photons, 10% neutrinos, 12% atoms https://arxiv.org/pdf/1907.10625.pdf Adam Riess
6. "Implemented in October 2013, Japan's new secrecy law will incriminate and imprison any Japanese for speaking of their illnesses, deaths and losses from Fukushima's ongoing radiation release.
7. APS Viewpoint: ASγ Arrays in Tibet see 100 GeV gammas from the Crab Nebula and **now 100 TeV!** HE electrons hit ambient photons through inverse-Compton scattering producing HE gammas. We now have 200 sources of TeV gamma radiation.
8. The Boulder Flatirons are hardened sand from wear-down of the ancestral Rocky Mountain Range that existed 280 Mℓa and then were pushed up by the current Rocky mountains 65 Mℓa.
9. Since the 1970's, we are in a new historical period in which everything is transformed and corrupted by the **neoliberal tools of financialization, deregulation and austerity:** laissez-faire economic liberalism, free market capitalism, privatization, reductions in government spending in order to increase the role of the private sector in the economy and society. Within this new nexus of power, anti-democratic and fascist principles have become normalized.
10. Rabbinic writings state that the Oral Torah and writings of Moses date to about 1300 BCE – **but** the world's first alphabet hadn't yet been invented. Archeology dates usage of P-H (paleo-phoenician-hebrew) for writing the Hebrew language to the 10th century BCE (not known in the Babylonian exile).
11. "Ask not about things which, if made plain to you, may cause you trouble." Quran (e.g., "Matrix moments, implications revealed by Trump's success).

12. Jeudi Noir. Black Thursday Like Trump"On the morning of July 16, 1942, four thousand French police officers descended on the Marais and other Jewish districts in Paris with orders to seize twenty-seven thousand Jewish

immigrants from Germany, Austria, Poland, the Soviet Union, and Czechoslovakia."

Physics Notes June 2019

Dave Peterson, June 4, 2019 – 7/4/19

1. In a first, LOFAR { Low-Frequency Array radio telescope network} has seen magnetic fields in a 10 M- ℓyr gap connecting galaxy clusters Abel 0399 and 0401. It could be that galaxies and the cosmic web are magnetized.
2. **Hubble Tension** suggests "physics beyond ΛCDM." CMB data do not measure" b and c parameters directly but rather $\Omega_b H_o^2(1+z)^3$ and $\Omega_c H_o^2(1+z)^3$ where z~ 1089 [1906.03947]. If somehow pressure p =wϵ then scaling is $(1+z)^{3(1+w)}$ and w~ -0.0108 solves problems. But w<0 implies that DM is not particles (and no DM particles have been seen). Its nature is unspecified but could imply Einstein-Cartan theory of gravity.
3. **Bohr Wrong:** NATURE "To catch and reverse a quantum jump mid-flight." The experimental results demonstrate that the evolution of each completed **jump is continuous, coherent and deterministic.** We exploit these features, using real-time monitoring and feedback, to catch and reverse quantum jumps mid-flight."
4. Astronomers found two new Earth-like planets, 12.5 lya in the habitable zone of an old red dwarf star. Each has 1.1 $M_\oplus$ minimum mass, orbiting at periods of 4.91 and 11.4 d, respectively. **Teegarden b** has actually scored the highest Earth Similarity Index (ESI) ever.
5. There are > 14 new 2019 gravitational wave detections! (April, May, June 2019) – but all are listed as preliminary reports still under evaluation.
6. **Phonons**: PRX: "Resolving **Phonon Fock States** in a Multimode Cavity with a Double-Slit Qubit." {Counting Phonons One by One}. JILA/CU 6/20/19 Superconducting device enables the detection of single quanta of sound, a step towards using them in quantum technologies--- they have some advantages—such as a long lifetime and a short wavelength compact size at the MHz frequencies where many quantum circuits operate.
7. Birds are Smart, and their forebrain can have 3 times the neuron density of primates. They evolved from warm-blooded agile theropod dinosaurs and are still dinosaurs. Now, "all birds are basically built the same way. It's what evolutionary biologists call a design restraint."
8. The formation of the moon via just "giant impactor" model doesn't quite work (similar composition of both Earth and Moon). A new model is "Synestia" also from impact but with extensive decades long baking and mixing in a hot wide toroidal cloud of rock vapor – the moon forms in the torus and Earth in center. [Sci.Am. July 2019 p 70]
9. The highest-energy photons ever seen hail from the Crab Nebula. An experiment in Tibet spotted photons with over 100 trillion electron volts of energy. {n=24} "AS-gamma" uses nearly 600 particle detectors spread across an area of more than 65,000 square meters in Tibet.
10. Sonic Hawking Radiation in a sonic horizon from accelerating a fluid of rubidium-87 atoms to supersonic speed. Phonon pairs at the horizon allow one to get swept with the fluid with featureless radiation (Jeff Steinhauer, Haifa)

The mathematical foundations of QFT are shaky or non-existent. Things like operator-valued distributions make no sense to mathematicians. The interacting QFT

Physics Notes May 2019

Dave Peterson, 5/2/19 - 5/22/19

1. https://www.nature.com/articles/s41586-019-1136-0 Analysis of the **kilonova** that accompanied **GW170817** identified delayed outflows from a remnant accretion disk formed around the newly born black hole as the dominant source of heavy r-process material from that event. Similar accretion disks are expected to form in collapsars.
2. **Collapsar accretion disks** (around their black holes) yield sufficient r-process elements to explain observed abundances in the Universe. Although these supernovae are rarer than neutron-star mergers, the larger amount of material ejected per event compensates for the lower rate of occurrence. We calculate that collapsars may supply more than 80 per cent of the r-process content of the Universe.
3. Stanley Prusiner: New Research "shows beyond a shadow of a doubt that amyloid beta and tau are both **prions**, and that Alzheimer's disease is a double-prion disorder in which these two rogue proteins together destroy the brain."
4. WITTEN: "I'd say that string/M theory **is the only really interesting direction we** have for going beyond the established framework of physics… " Loop Quantum Gravity (as another possible route**)—"those are just words**. There aren't any other routes." ‼
5. Why the General increase in IQ: After 1920's in America, iodine added to salt boosted IQ by 15 points (part of the "Flynn" effect [Discover]). Also, poverty reduces brainpower and concentration (equivalent to 13 point drop in IQ).
6. Mikhail Gorbachev claimed that the **Chernobyl explosion** in 1986 was "perhaps the real cause of the 1991 collapse of the Soviet Union". Like Fukushima, the economic impact was hundreds of billions of dollars. Aided by Glasnost, Soviet citizens questioned state infallibility and realized that their government and industries were startlingly incompetent leading to the death of the Soviet Union in 1991. Finally, since the Fukushima disaster of 2011, the worldwide nuclear industry is now in serious decline. And, surprisingly, a rapid Japanese switch to safe power kept their carbon emissions at "normal" levels—we don't really need the nuclear industry.
7. Theoretically, one can store 455 exabytes of data in a single gram of DNA ‼ And it is now easy to read millions of DNA sequences at the same time.
8. [1812.08336- Ohio State] claims to calculate ϵ_o from vacuum fluctuation e^+e^- pairs assuming them to form a bound state of zero angular momentum for a short time. So e_o and c are properties of the vacuum polarization from electric fields. BUT, gravity waves and neutrinos also travel at c making that speed more fundamental than just QED [and, is there adequate time for forming positronium bound states?].
9. Short: During my 30 years at StorageTek, I wrote an average of 130 reports per year. "The moment we stop believing in the Financial market, it ceases to exist." Around 1860 the population of Marshall was greater than Boulder. Einstein's 1905 photon paper counted **against** him in the physics community for nearly 20 years! He was continually pressured to reject it.

Physics Notes April 2019

Dave Peterson, 4/2/19 -4/29/19

1. Nelson (1985) defined (the dogma of) "**naturalness**" in terms of simplicity, beauty, unification, cosmological principle, insulation (succeeding scales are insulated from one another) -- It was used to show problems with the SM and that SUSY could solve them. Unnatural was quadratic divergences of the Higgs mass, fine-tuning, renormalization group instability, and the ratio of two very different masses like those of the light and heavy quarks. After LHC non-results, naturalness is on the wane. https://arxiv.org/pdf/1904.01450.pdf
2. **Dark matter experiment finds no evidence of axions**
In its first run, MIT's ABRACADABRA detects no signal of the hypothetical dark matter particle within a specific mass range 0.31 to 8.3 neV detecting signals less than 20 atto-Tesla.
3. For meals: "*Blessed are the Universal Works of Nature that enable bread to emerge from the Earth (biosphere).*"
4. ScienceNews NEW Here are 5 RNA categories that are stepping out of DNA's shadow; these molecules play crucial roles in human health and disease. There are now 25,000 known genes with instructions for **noncoding RNAs** in the human genome (and many more to come) – versus 21,000 coding genes for making proteins: LIST 1. 18,000 lncRNAs "long noncoding RNAs" (pronounced "link RNAs"). 2. MicroRNAs barely more than 20 RNA units, or bases long. 3. Transfer RNAs, or tRNAs (making protein). 4. SINEs for short interspersed nuclear elements make RNA copies of themselves. 5. piwi-interacting RNAs, or piRNAs (pronounced "pie RNAs").
5. **EHT on M87*** : https://iopscience.iop.org/issue/2041-8205/875/1 EHT giga-black hole results April 10, 2019 are: Crecsent diameter ~ 42 μas, "angular gravitational radius GM/Dc^2 = 3.8 μas, D = distance. Mass 6.5 Gsuns by stellar dynamics and now by imaging. Shadow angular diameter 38 μas. The photon capture radius is 2.6 x Rc ($2MG/c^2$). Ring width < 20 μas
6. **New Human: Homo luzonensis**, whose teeth and bones were discovered in a cave on the island of Luzon in the Philippines. The remains represent a new species, scientists concluded in a report published Wednesday in the journal Nature.{ 67 kya, < 4' tall,
7. Climate change affects the change in O_2/N_2 over short durations. But over millions of years, it is plate tectonics such as India slamming into lower Asia about 50 Mya or Antarctica accumulating ice 35 Mya. (source foraminifera shells and N15/14 ratios).
8. New measurements from NASA's Hubble Space Telescope confirm that the Universe is expanding about 9% faster than expected based on its trajectory seen shortly after the big bang, -- and local H is > Planck CMB H.

A Black Hole Merger Every Week? LIGO And Virgo Are Back! (mergers on 4/8, 12, 21, 24,26 but only 4/8 reported publicly by Ligo).
https://www.youtube.com/watch?v=4GfjB5vzq3g
https://www.ligo.caltech.edu/ for official notifications.

Since LIGO and Virgo resumed operations this April after undergoing upgrades to improve their sensitivity, they have recorded five candidate gravitational wave events (one of which may be a kilonova) — as many as were spotted in LIGO's first two years of observing. https://twitter.com/LIGO/status/1123336957193523204\. The latest detection, on April 26, may even be from a never-before-seen event: the collision of a neutron star with a black hole. Provisionally labeled S190426c,

Book Report The Shape of a Life Shing-Tung Yau and Steve Nadis, Yale, 2019

Physics Notes March 2019

Dave Peterson, 3/5/19- 4/1/19

Recent Summary:

1. A vast lake lies 1.5 km underneath Mars's southern pole and stretches 20 km across. The MARSIS instrument on the Mars Express spacecraft sent out radar pulses that penetrated the surface and ice caps on Mars, and measured the radio waves when they came back to the spacecraft (data from 2012-2015+ years of verification).
2. Muon-telescopes (e.g., "G3MT") note a relatively precise reduction of atmospheric anti-muons due to the potentials in thunderstorms. A thunderstorm studied in December 2014 derived a peak record-breaking electric potential of **1.3 GV** --ten times higher than previous records.
3. EHT on Sgr A* mega-black-hole: The current highest-resolution (approximately 30 μas) measurement, made at a wavelength of 1.3 mm, indicated an overall angular size for the source of 50μas. As of April 2017, there have been direct radio images taken of Sagittarius A* with the Event Horizon Telescope, but the data is still being processed, and images have yet to be released.
4. the GRAVITY Collaboration just performed a black hole test using the light emitted from a star orbiting Sagittarius A* -- Einstein GRT still OK!
5. Nearly spherical Fermi-bubbles of galactic size lie just above and below the plane of our galaxy and emit gamma rays (of unknown source cause). Now we also see x-ray "chimneys" about 300 ℓy wide "funneling through the galaxy's center into the gamma-ray bubbles." (Nature, March, 2019).
6. **On-demand** Semiconductor Source of Entangled Photons Which Simultaneously Has High Fidelity, Efficiency, and Indistinguishability, Hui Wang, et al., https://arxiv.org/pdf/1903.06071.pdf & Phys. Rev. Lett. 122, 113602
7. Nuclear astrophysics DFT, [1104.1194]: **Neutron star** outer crust $10^4 \leq \rho \leq 10^{11}$ g/cc surrounds an inner crust (up to 10^{14}) outer core to 5×10^{14} (homogeneous liquid of n, e, p and μ's) and inner core up to $10\rho_o$ (where nuclear $\rho_o \sim 1/6$ $fm^3 = 1/(1.8fm)^3$)—but neutron diameter ~ 1.6 fm and p-dia ~ 1.7fm (so there is no room for inter-penetration if n's and p's are "particles" in nuclei).
8. Physics Today March 2019: New data seems to show that white dwarf stars must crystallize as they cool down [statistical survey by European Space Agency's Gaia satellite observatory of 1.1 billion stars]. This causes a pile-up bump in cooling rate of the WD's—and it is seen. WD's are oxygen and carbon with oxygen settling to the core and solidifying into a body-centered cubic metal.
9. Physics World: Particle physicists at CERN have measured charge-parity (CP) violation in the D0 meson for the first time (5.3 σ > 5.0)– It is the first time that CP violation has been seen in **charm** mesons and opens up the possibility of searching for physics beyond the Standard Model.
10. Hidden fact: Armstrong and Aldrin were almost stuck on the moon because of a broken circuit breaker to their rocket engine—and for many hours, NASA didn't know how to fix it. Fortunately, Buzz had a felt-tip pen that he inserted in a hole to finally activate the circuit.

11. "Living with the New SI" (physics.aps.org) on May 20, 2019 we will have exact physical constants: h,e,k_B, N_A, m_e -- and old losers are: kg, V, Ω , μ_o, ϵ_o, T_{twp} (triple point of water) – no more artifacts for standards. NIST is trying to put itself out of business.

Physics Notes January – February, 2019

Dave Peterson, 1/5/19- March 3,2019
[Absent most of January, Hospital]

1. Massive black holes at galaxy centers are supposed to have many smaller ones nearby. "We have now detected a **dozen black holes at the center** of our Milky Way galaxy, with as many as 10,000 projected [Nature].
2. APS: [4 meter Blanco telescope in Chile]: The **Dark Energy Survey** (DES) has completed a six-year observing run. It shows that "Weak lensing becomes a high-precision survey science" competitive with CMB (but CMB is at age 380,000 years ABB while DES sees billions of years later). ΛCDM is still an accurate description.
3. Some physicists argue that gauge symmetry is not really a fundamental feature of nature but merely a technical redundancy. Zee: gauge symmetry is strictly speaking not a symmetry but a redundancy in description. It is not physical nor observable (Matthew Schwartz), it is completely garbage (Nima Arkani-Hamed). [1901.10420].
4. OxyContin is stronger than Morphine but was marketed as weaker and more like Hydrocodone (a lie). Aggressive marketing of OxyContin is blamed for fostering a national crisis that has resulted in **200,000 overdose** deaths related to prescription opioids since 1999.

5. https://arxiv.org/pdf/1902.05108.pdf, Eliahu Cohen, Marina Cortes, Avshalom Elitzur, and **Lee Smolin,** **"**Realism and causality I: Pilot wave and **retrocausal** models as possible facilitators," suggest that a retrocausal version of the pilot wave theory, in which the particle is guided by a combination of advanced and retarded waves, might account for quantum physics with less damage to intuition. Refers twice to Cramer and also Kastner and Aharonov. Cramer's confirmation waves are similar to Aharonov's destiny vector for corpuscle final position. Space-time is emergent from quantum collapses.

6. [The Inner Lives of Neutron Stars, Sci. Am. March 2019 p 24]. Neutron stars have kilometer-deep outer "crust" made of atomic nuclei arranged in a crystal structure with electrons and neutrons between them." Nuclei with too many neutrons spill out into just neutrons with no nuclei—this might become a superfluid and possibly a quark superfluid with bound Cooper pairs. **"Scientists are fairly sure neutrons in the crust" are paired.** "Glitches" may occur when the frictionless superfluid rotates faster than the star – mismatch catch-up "earthquake." Masses and radii of pulsars can be measured by the NICER instrument on the space station (x-ray photons timing and energy). Strange hyperons composition is now limited to 10%
7. **REAL REPRESENTATION of C**. a + bi = rexp(iϕ) positive rotation ccw. [let c=cos, s=sin]: The rotation matrix in 2D is R = [c –s; s c] = c[1,0; 0,1] + s[0 -1; +1 0] = c1+si. So **i → [0 -1; +1 0],** and i i = -1I.
8. https://arxiv.org/pdf/1902.11196.pdf, An Illustrated History of Black Hole Imaging : Personal Recollections (1972-2002) Jean-Pierre Luminet (to supplement new

book: Einstein's Shadow by Seth Fletcher (on event horizon telescope for seeing Sag A* mega BH).

Physics Notes December 2018

Dave Peterson 12/6/18 -1/3/19

1. In 2017 entanglement swapping was achieved in a quantum network entangling two photons over a distance of 100 km (Optica). Since initial biphoton production by SPDC is inefficient, a better source is needed. This could be identical particles simply entangled via their indistinguishability using beam splitters. 1812.02141

2. Neutron star mergers create the heaviest chemical elements: Spitzer mid-IR NS merger GW180817 now seems to include the third peak of the r-process (Ba, Pr, Nd, Eu, Os, Ra, Pa, Th) – these heaviest elements were synthesized. Peak 1 is A = 70-88, Pk2 120-140, Lanthanides 139-180, 3^{rd} pk 180-200 – very red spectral energy distribution. (MN Dec 2018).

3. **The AdS/CFT conjecture** relates gravity in the wrong space-time dimension (5), with the wrong space-time curvature (AdS) to a quantum field theory that doesn't describe any known particles (N=4 SSYM). For the last twenty years there has been lots of speculation about the possibility of extending this to the real world cases, but this hasn't worked out. There's no known dual QFT to our gravity…

4. Russia is poised to add a new hypersonic rocket and nuclear warhead to its arsenal, the Avangard capable of Mach 20! = 20x1080 km/hr. Dia Earth 6.37 Mm. Time to opposite on earth is $\pi R/v \sim 1$ hour! Very little time to respond.
5. Red and blue America aren't separated just by their cultural politics; they are separated by sharp differences in how their **economies** have developed over the past half-century. And those economic differences can, in turn, explain many of the cultural differences that so bedevil our political system. … the older industrial economy versus the newer **ideas** economy [finance, technology and electronics].

6. Far from being just a 2-level Quantum system the Qubit is a Unit Quaternion, also known as a Spinor. Therefore the Qubit is a 4-dimensional vector which traces a path on the surface of the unit 3-sphere, S^3. This is the meaning of the global phase (θ, ϕ, α). We go to the 3-D "Bloch sphere" by ignoring the alpha phase.
7. Scientific American Jan 2019: Humans evolved to exercise, we require high levels of physical activity in order to be healthy. In contrast, apes are lazy and thrive that way. We are built to move, and our exercise/ambition took us out of Africa. **Our power output is 4 times** that of Chimps. "Weight loss is the one health benefit it largely fails to deliver." (Hadza hunter/gatherers and sedentary Westerners have the same energy expenditure.
8. Sci Am: The Particle Code: there is a new discipline "amplitudeologists" (e.g., Amplituhedron) greatly simplifies and replaces complex Feynman diagrams – replaces higher order loops with periods and logarithms and can do up to seven loops (α^7).
9. "Evolution has a great memory but no plans." Example is Homo naledi (0.335-0.236 mya in South Africa) having near human body but a very small brain size with no further growth over 100,000 years – it was adequate.

10. “Every day, the U.S. nuclear early warning system is triggered by some event or another, mostly civilian and military rocket launches by one or more of a dozen countries with ballistic missiles.
11. It took a century to finally pass an anti-lynching law in America (just now) -- because lynching had remained a powerful terrorist tool to maintain white supremacy.

12. Kuiper Belt object encountered: New Horizons is expected to collect and store around 7 gigabytes of data during the flyby sequence, then transmit the information back to Earth at a speed of about **1,000 bits** per second off-and-on over the next 20 months. (little 15 watt transmitter 6 light hours away).

13. T.D. Lee’s two laws of physicists: “Without experimentalists, theorists tend to drift. Without theorists, experimentalists tend to falter.”
14. Feynman: Sometimes you'll hear that light is made of photons. What that means is that when light is absorbed or emitted, the energy in the wave comes in lumps. "Electrons acts like wave, no they don't exactly. They act like particles, no they don't exactly.”

New Notes:

More than 90 percent of the pyrite on Earth is formed by microbiological processes. Bacteria also catalyze pyrite’s oxidation and breakdown.

My Amazon Review: HalfBreed, An amazing and wonderful book by David Halaas (d 8/ 2019)
This book was very rich in Indian history representing the last half of the 19^{th} century. Since I was a kid, I’ve had an attachment to Indian culture – especially living in and being a part of Nature as opposed to our highly artificial culture. As a “halfbreed” with exceptional memory, **George Bent** had a unique ability to communicate that culture to us. His own life is as remarkable as the book cover indicates. Having grown up in Kansas and Colorado, his history overlaps with my experience—removed by a century in time. And the Cheyenne are one of the more interesting tribes. I underlined a lot of the book. This is actually my second reading, and I enjoyed it as much as the first. A lot of people would benefit from knowing about and reading this book for themselves.

Physics Notes November 2018

Dave Peterson, 11/4/18 -12/5/18

1. Trans-galactic streamers are feeding the most luminous quasar galaxy in the universe emitting IR light like 350 t-suns! (vs usual 0.1 t/ MW type). W2246-0526 at 12.4 blya is cannibalizing at least 3 neighboring galaxies. Its central MBH may be 4 b suns. [ScienceDaily].
2. First noted in 2014: Quasars can turn on and off during human lifetimes! Mechanisms are not yet clear.
3. Light-matter **photon-phonon entanglement** between the vibrational motion of two silicon **optomechanical** oscillators each having 10 Billion Atoms , 11/ 29, 2018 showing 4-σ Bell inequality violations. The two **macroscopic** ‼ mechanical resonators are μm long separated by 20 cm. A blue pulse drives the vibration, and a red pulse detects the phonons. https://physics.aps.org/synopsis-for/10.1103/PhysRevLett.121.220404, https://arxiv.org/pdf/1806.10615.pdf

New Notes:

Adequacy of Popper Falsification? : An example of this is given by the discovery of the planet Neptune: when the motion of **Uranus was found not to match the predictions of Newton's laws (falsified?),** the theory "There are seven planets in the solar system" was rejected, and not Newton's laws themselves. … Paul Feyerabend ultimately rejected any prescriptive methodology, and argued that the only universal method characterising scientific progress was anything goes.

Hermann Weyl said of Galois final testament (Group Theory):
"This letter, if judged by the novelty and profundity of ideas it contains, is perhaps the most substantial piece of writing in the whole literature of mankind."

Physics Notes October, 2018

Dave Peterson 10/5/18 – 11/3/18

1. "rapid, far-reaching and unprecedented changes in all aspects of society" is needed to keep global temperatures from rising more than 1.5 degrees. But there is severe economic impact for a solid fix and an impact even for likely half-measures. So republicans have decided to avoid the issue altogether.
2. Jefferson Lab e+p→n pion data shows that the momentum contribution at GeV experimental energies is: 54% from valence quarks, 16% from sea quarks, and 30% from gluons. The previous belief was that only 10% was due to gluons in the pion [APS].
3. [1810.04341] Vacuum Fluctuations low order calculations give the permittivity of the vacuum (2.8% above actual ϵ) IF e^-e+ pairs act as bound states allowing VF's to act as harmonic oscillators (how likely is that?).
4. https://arxiv.org/pdf/1810.04823.pdf **12-photon entanglement** and scalable scattershot boson sampling with optimal entangled-photon pairs from parametric down-conversion, China, GHZ state $(|H\rangle^{12}+|V\rangle^{12})/\sqrt{2}$. 12 photon coincidences are 1/hour for a 2 MHz input.
5. Hubble himself never knew he had discovered the expansion of the universe (1953). Gamow gave him credit (1952) but Lemaitre was first (1927). Peebles said "Physical scientists have a healthy attitude toward the history of their subject: by and large we ignore it." Slipher should be more recognized.
6. Woit: LHC Run 2 13 TeV from 2015- end on 10/24/18. Now shut down until Run 3 in 2021. Data processing is only up to 2016 so far.
7. **UK Renewables succeeding too well**: power demand has a typical summer daytime peak of around 30 GW, but at night can fall to 17 GW—and demand is falling. At present there is around 42 GW of renewable capacity installed, supplying nearly 32% of UK electricity, and more is on the way; by 2020 there should be 46 GW and by around 2027 maybe 60 GW, on current plans. This is mainly on-shore and off-shore wind turban power with PV solar being minor. But, its Hinkley Point C nuclear plant may start up by 2027 with an excess 3.2 GW capacity—what to do with regular excesses. Germany's problem may be too much solar.
8. https://physics.aps.org/articles/v11/109 Two teams demonstrated that they can count the number of quantized vibrations, or phonon bosons in cold mechanical oscillators by measuring the energy in the vibrations. Trial center of mass vibrating membrane connected to a superconducting qubit can control **phonon number Fock states!** There can be single phonon states and ground states are zero phonon states.
9. PW: Blue supergiants of mass > 50 suns go supernovae but should not do so according to standard theory. Upon the infall from the first explosion, intense pressure might induce a phase transition from neutrons to quark-gluon plasma

(QGP) causing a stronger explosion. In this two-stage successful process, the final remnant is still a neutron star.

10. LIGO Analysis: https://www.sott.net/article/399642-An-illusion-Grave-doubts-over-LIGOs-discovery-of-gravitational-waves

"**The Meaning of Life is to give Life Meaning.**" Victor Frankl

Physics Notes September 2018

Dave Peterson 9/4/18- 10/1/18

1. Record longest slow-spin pulsar 23.5 sec (LOFAR LOw Frequency ARray radio) PSR J0250+5854—versus fasted at 1.4 ms.
2. Higgs Factory Race: China now intends to bypass the 3 TeV e+e- CLC and instead produce a **China Electron Positron Collider (CEPC)** [100 km circle, 240 GeV, $6 bn, millions of Higgs]
3. deBroglie/Bohm: NonLocality has its origin in the guiding wave propagating in multidimensional **configuration** space! John Bell [Quora, 9/17]. Bohm trajectories are averages of an ensemble of Feynman paths [Hiley]. [Note: Woit's last book on quantum mechanics made no mention of probability.]
4. A recount of human genes ups the number to at least 46,831 – almost double because non-coding genes make RNA beyond that needed to make proteins.
5. The Roman Empire was home to the longest conflict in human history, the Roman-Persian Wars. These wars went on for an estimated **721 years** – from 66 bce.
6. The reason Trump remains is simple: He occupies the best political ground, namely the meeting point of **three reactionary forces** in American life.
7. The **tycoons** business elites, White Nationalists, and White Evangelical Protestants.
8. Earth Oxygen levels 1.87 bya were $1/1000^{th}$ of present levels. !
9. PRL Nearly 10 years of Fermi telescope images show unexpected changes in the numbers and energies of gamma-ray photons coming from the Sun. The team also detected nine photons—all from the equator—with energies exceeding 100 GeV, the first detections from the Sun at such high energy.
10. The $p=h/\lambda$ deBroglie wave is a result of Lorentz transformations of an $\omega = E/\hbar$ intrinsic intrinsic rest-mass particle vibration from effective clock desynchronization and of course varies with observer speed, v.

Physics Notes August 2018

Dave Peterson, August 3, 2018 --- 9/4/18

1. Feynman Nobel: Dirac Hole negative energy sea theory is more complicated but **equivalent** to letting positrons run backward in time .
2. The gauge field A^μ is an connection that "lives in the algebra of the gauge group" which is little **u**(1)= R ! "It is better to say that free electromagnetism is the theory of a **u**(1) gauge symmetry" (the **Lie algebra** of the generators of the continuous Lie group) rather than U(1). The Gauge Principle is the notion that a global symmetry should continue to hold at the local level.
3. Glutamate is a chemical that nerve cells use to send signals to other cells (only realized after 1970). It is now by a wide margin the most abundant neurotransmitter in the vertebrate nervous system.

4. Status: there is no compelling reason to prefer **quantising gravity** over developing QFT in curved space-time, but neither is easy and the Physics community is not yet convinced by any of the proposals. Lubos says: Gravity has to be subject to quantum mechanics because everything else is quantum, too (not acceptable).
5. String theory permits a "landscape" of possible universes, surrounded by a "swampland" of logically inconsistent universes. **In all** of the simple, viable stringy universes physicists have studied, the density of dark energy is either diminishing or has a stable negative value, unlike our universe, which appears to have a stable positive value !
6. The **New Horizons** spacecraft has spotted an ultraviolet glow that seems to emanate from near the edge of the solar system coming from a long-sought wall of hydrogen that represents where the sun's influence wanes (like Voyager punching into interstellar space).
7. Physics Today, Aug 2018 **Megadrought**. About 1150 C.E. occurred the greatest drought in American history by far but a later next worst near 1280 (decades long) may have hit Puebloans at Mesa Verde and Chaco Canyon
8. DNA on bones from a 13 year old girl 90,000 years ago showed a Neanderthal mother and Denisovan father! All this intermingling suggests that all are Homo sapiens species [The Denisova Cave is named after Denis, a Russian hermit who lived there in the 18th century].
9. Black Holes might be string-theory fuzzballs with no horizon and no singularity and no effective burn-up – But, not even string theory can handle the messiness of realistic fuzzballs.

New Notes:

Jimmy Carter: "Our Supreme Court has now said, 'unlimited money in politics.' It seems like a violation of the principles of democracy.
"It violates the essence of what made America a great country in its political system," the oldest living Democratic ex-president replied. **"Now it's just an oligarchy,** with unlimited political bribery being the essence of getting the nominations for president or to elect the president. And the same thing applies to governors and U.S. senators and congress members."

Physics Notes July 2018

Dave Peterson, 6/30/18 - 7/29/18

DiHiggs: Can the Higgs do h → hh? (gluons can, g→gg). This is still an open question that might be resolved in future LHC Run3. We do not yet know if the Lagrangian allows for self interaction of the Higgs.

1. In June 2014, IceCube saw light from a charged lepton that deposited an extreme energy of 2.6 PeV in the detector. The early belief was a 10 PeV muon neutrino. But now, 4 years later, calculations indicate that it might have been a 100 PeV tau neutrino.
2. **Blazar** 0506+056 HE **Neutrinos** (4 Glya, 9/22/17, z=0.336): IceCube South Pole agrees with gamma telescope observation indicating **the same source** (also radio and optical): 300 **TeV** nus and γ's exceeding 400 GeV (n=54 HE nus seen so far). Lorentz invariance nearly perfect (1807.05155)

3. Heisenberg's 1932 Nobel prize in physics for QM emphasized the discovery of the "allotropic forms of hydrogen." **Parahydrogen** (proton spins opposite) is in a lower energy state than is orthohydrogen, but room temperature yields 75% orthohydrogen. (I had never heard this before).

4. Remember path-integral people including **non-classical paths** through slits and back out and in again –well, just done. Microwaves λ =5cm, 3 slits w=10cm and separation 3 cm, source and detector 1.25 m to slits. Deviation 6% seen from the superposition principle (New Journal of Physics) PW Sinha.
5. Danish, Boulder Weekly July 19, 2018: Rocky Flats is supposed to be a national wildlife sanctuary **not** a national recreation area open to the public. It should remain closed because of residual contamination (Americium , U233, U234, Pu), lack of trust, along with deeply buried Pu.
6. Schwinger's careful use of source theory can avoid the need for renormalization, zero point energy, superpositions, and particles --except as final quanta of fields) – quantized field replaces classical duality.
7. CERN data up to 13 TeV now shows Higgs coupling to b-bbar and t-tbar quarks with nearly adequate statistical significance and "give a strong indication that the Higgs boson has a key role in the large value of the top quark mass" and all four primary modes of Higgs production have now been observed at the LHC.
8. **Why Riemann Zetas?:** 1512.09265 Periods and Feynman Amplitudes. Magnetic moment sum of Feynman amplitudes for two-loop diagrams give: 197/144 + ζ(2)/2 – 3ζ(2)ln2 +3ζ(3)/4 ~ -0.3284, ζ(2) = π2/6. ζ(3) = 1.20205. (proven irrational in 1978), ζ(4) = π4/90, ζ(-1) = -1/12. Z(1) = ∞ , ζ(5) = 1.03692.

New Notes:
---Rovelli's Relational QM (interaction between system and observer) should include TI as a special case. Data is slowly accumulating to show the existence of sterile neutrinos.

"I keep six honest serving-men (They taught me all I knew); Their names are What and Why and When; And How and Where and Who. Rudyard Kipling,

"Talent is hitting the target nobody else can hit, while genius is hitting the target nobody else can see." Schopenhauer

Dulong and Petit formulated their limiting law (~1819) at a time when Lavoisier's **caloric** prevailed and Dalton atomic theory was new. 1807.02270

Physics Notes June 2018

Dave Peterson, 5/28/18 – 6/28/18

1. **LIGO** [1805.11579] GW170817 [**BNS** N+N] masses 1.2-1.6 suns and spins 0.5 and 0.6. [1805.11581] adds Radii ~ 12 km, pressure near 2x nuclear saturation. (Sat = $\rho = 2.8\times10^{14}$ g/cc).
2. Max seen NStar-**spin,** f~716 Hz (PSR J1748-2446ad)—well below the current predictions of theoretical equations of state (breakup frequency > 1200 Hz) [1805.11277]
3. About 66 million years ago, the Chicxulub asteroid impact set off **100,000 years of global** warming (Oxygen isotope ratios).
4. After years of U.S. dominance, China is closing the science gap and will surpass the United States in spending on scientific research by the end of this year, according to the National Science Board.
5. NOvA/Fermilab sees muon antineutrinos oscillating into nu_e's (500 mile through Earth path, count 18 when 5 expected).
6. The Simons Foundation's math/physics fundings are now comparable to those of NSF (hundreds of $millions).
7. Since 1965, Linus Pauling published 25 article on his close-packed spheron model of atomic nuclie (though not accepted in the physics community).

8. 2012: The physical explanation of the covalent bonding mechanism is **still being debated** (e.g., Lewis theory~ enhanced middle electron density vs

reduced kinetic energy from delocalization of valence electron motion and contraction of atom orbitals).

9. Humans have ~20k genes and one of them has ~200k base pairs. A change of **one letter** from G to an A in that gene 8 kya produced white people (Skin Color gene SLC24A5 – the problem of **race**!). National Geographic.
10. It is possible that the Type-2 diabetes epidemic is largely due to increased prevalence of white titanium dioxide pigments that can lodge in the pancreas.
11. Muon g2 requires not just Dirac and QED but also weak interactions (W's+v and Z) and hadron calculati0ns (from electron-positron collision data). Presently 12,672 Feynman diagrams have been calculated down to better than 1 ppm. A small 3σ difference seems to exist between theory and experiment.

New Notes
[1805.11501]: Blandford, Cosmic Rays up to 100 EeV ‼ (GZK curoff photo-pion production on CMB). Ankle ~ 3 PeV 90% protons

About 66 million years ago, the Chicxulub asteroid impact set off **100,000 years of global** warming, an analysis of oxygen in fish fossils suggests. It's not surprising that the climate heated up after the collision, which left a 200-kilometer-wide crater centered around what's now Chicxulub, Mexico,… The ratio of heavier oxygen to lighter decreased by about 1 percent in the fish bits collected after the impact compared with those pieces from before the impact, the team found. That change translates to an increase in seawater temperature of about 5 degrees Celsius — a substantial amount.

The world's smallest **atomic clock chip** is 4 x 3.5 x 1 cm in size, weighs 35 grams, and consumes only 115 mW of power. It is accurate to 501 ns per day, or approximately one second in 5001 years. The Cs-133 chip keeps time with an…

Physics Notes May 2018

Dave Peterson, 4/23/18 - 5/27/18

1. Dinosaurs from 230-200 Mya were unable to compete against pseudosuchians (false crocodiles) until the prolonged lava flow from the breakup of Pangea (end Triassic).
2. ***Multimessenger events:*** #1 SN 1987 A in neutrinos, γ, X, then optical. #2 Texas source 9/22/17 IceCube neutrino 0.1EeV! then Swift x-rays 9/26, Fermi γ 9/28, then 9/29 optical Blazar source 50x increased brightness, then radio. #3. 8/17/17 **Ligo**-Virgo n+n kilonova gravity! Then γ, optical, IR (Sci Am May 2018).
3. **Rovelli:** Three major empirical results have marked recent fundamental physics: gravitational waves, the Higgs, and the absence of supersymmetry at LHC. All three are confirmations of old physics and disconfirmations of widespread speculation" Do not speculate so freely. Nature is snubbing current methodologies.
4. There are No mountains at Yellowstone (they all fell into the supervolcano),
5. APS: OPERA has seen 10 tau neutrinos produced from an input of muon neutrinos [n=19,505 detected (2008-2015)]. Out of $2x10^{20}$ protons on target.
6. **Conversions: 70 lumens/ LED-Watt, old watts/5 for new LED watts.**
7. Quantum entanglement has crossed over from the minuscule to the very small (1-15 μm) drum heads and silicon beams – the first time for macroscopic structures. The aluminum sheets have a trillion atoms. The drum heads interact with microwaves and get in synch (for 30 minutes!) and the beams with IR (Nature vol 556, April 26).
8. The arrest of Mikhail Khodorkovsky, the head of the Yukos oil company, in October 2003, was a key turning point in modern Russian history. From being

one of the world's richest and most powerful men, Khodorkovsky became Putin's prisoner.

9. **V-A:** "vector – axialVector" charged-current weak interaction from Marshak and Sudarshan, ½(Cv+ γ_5 Ca), neutron beta decay Ca/Cv = -1.26 but Λ β decay -0.72. Form g_v -g_a γ^5. But neutrinos are 50-50%, and electrons are -0.054 to -0.5 (mainly axial).
10. California's gross domestic product surpassed $2.7 trillion from 2016 to 2017, making the U.S. state the world's fifth-largest economy, bigger than that of even the United Kingdom
11. Whatever else Trumpism may be, it is the systematic organization of resentment against outgroups. Trump's record is rich in dehumanization. Because we are **inherently predisposed toward stereotyping,** we are particularly vulnerable to propaganda.
12. Quantum Computing is now solving real world problems using 2-6 qubits. IBM offers 5-16 qubit cloud computing with 50k users! It just calculated the binding energy of the deuteron mainly using 2-3 qubits with 2-3% accuracy.

Comedian: "He wants to give teachers guns, and I support that because then they can sell them for things they need like supplies.

YouTube by Robert Spekkens, "The Riddle of the Quantum Sphinx", example of a ``Wrong Category:" Egyptian hieroglyphics thought to be ideograms (wrong), now QM psi is real (he says instead that ψ is state of knowledge).

Physics Notes for April 2018

Dave Peterson, 3/23/18- 4/22/18 (32 pages).

1. Feynman: could always look at something the way a child does. He sees things with curiosity and wonder, finding something new …and always thinking in pictures. After QED, he abandoned absorber theory—Lamb shift is a self interaction [Paul Halpern, book]. [But Wheeler re-considered it in 2003].

2. Wikipedia [the 'good cop' of the Internet] is an exception to the Chomsky propaganda model: By 2004, Wikipedia swore off advertising completely after its community of volunteers threatened to take their contributions and create a separate site
3. There is mounting evidence that string theory abhors deSitter vacua (cannot yield our universe).
4. "Observation of Entangled States of a Fully Controlled 20-Qubit System," PR-X, Innsbruck. The researchers were able to detect genuine multi-particle entanglement between all neighboring groups of three, four and five quantum bits
5. Researchers have created a Bose-Einstein condensate of light coupled with metal electrons, so-called surface plasmon polaritons on a gold nanorod array.
6. The view of "spooky action at a distance" by a philosopher (Joan Vaccaro) is that it is analogous to the 1500's puzzle of the retrograde motion of the outer planets. That motion wasn't fundamental but rather due to a wrong "fixed Earth" perspective.
7. (93% certainty) Between 17% -35% **of Americans are atheists**, with a "most credible indirect **estimate" of 26%.** 64M American Atheists -- ground our morals and values on reason and science. SCI AM April 2018 p77:

8. The first known superconductor in which two spin-3/2 quasiparticles form Cooper pairs has been created by physicists in the US and New Zealand. The unconventional superconductor is an alloy of yttrium, platinum and bismuth, which is normally a topological semimetal. 3/2 results from spin-orbit coupling and topology at a temperature of 0.8 K.
9. Our Moon Rotating https://apod.nasa.gov/apod/ap180318.html -- can't see this from Earth; front and Back are very different! Also see CRAB Pulsar. https://apod.nasa.gov/apod/ap180317.html

Joke: A Soviet citizen parked in front of the Kremlin. "A policeman rushed over to him and yelled, 'Are you crazy? This is where the whole government is.' No Problem, said the man. 'I have good locks on my car."

New Notes:

Common Taboos: Cannibalism, incest, bodily functions, murder, abortion, adultery, abnormal sex, suicide, corpses, eating carnivorous animals , food and drinks, inter-racial sex [sex between races Inter-religion Marriages].

There is little reason to believe that **instincts** that evolved to shape our survival in a hunter gatherer community would be useful in helping us triumph in a complicated world consisting of nation states. How about patriarchy?

Physics Notes for March 2018

Dave Peterson, 2/21/18 - 3/20/18

1. The history of physics cannot be well understood without appreciating the unbelievable antagonism between the Chew/Mandelstam/Gribov S-matrix camp, and the Weinberg/Glashow/Polyakov Field theory camp. **The two sides hated each other**, did not hire each other, and did not read each other. In the 1970s, S-matrix theory just plain died.
2. New big DM simulation: still too many dwarf galaxies and still peaked cusp distribution problem. Milgrom still does better.
3. New DNA analysis: We are more like other domesticated animals (we domesticated ourselves!). Neanderthal genes indicate that they were NOT domesticated. So our **socialization** made the difference.
4. The only remaining trait defining dinosaurs: The hole in the hip socket probably helped dinosaurs **position their legs underneath their bodies**, rather than splayed to the sides like a crocodile's legs.
5. We know that two entangled photons can act as a biphoton with joint detection wavelength $\lambda/2$. A recent test using two entangled **electrons** (from double ionization of H_2 molecule due to 400 eV photons) acts as one quasi-particle with narrower wavelength due to momentum k_1+k_2 effective addition interference (1607.07275)
6. Dark Matter and the Earliest Stars: observational signature from the very first stars in the universe formed ~ 180 million years after the Big Bang (a little over one percent of the current age of the universe).
7. There has been a 40% decline in absolute mobility from 1940 to 1980 (children earning less than their parents)
8. **Astronomers have discovered that all galaxies rotate once every billion years,** no matter how big they are… by using simple maths, you can show all galaxies of the same size have the same average interior density.
9. About 70,000 years ago, a small reddish star approached our solar system and gravitationally disturbed comets and asteroids. Astronomers have verified that the movement of some of these objects is still marked by that stellar encounter. [SD] Scholz's star came within 1 ly from us.

10. Sabine (on Afshordi): If the horizon of a black hole is obstructed by something like a firewall, then the horizon could potentially reflect gravitational waves. LIGO has seen that echo—but perhaps only $\geq 1\sigma$ so far.

Physics Notes February 2018

Dave Peterson 1/8/18- 2/10/18

Summary of Recent Notes:

1. “Using Gravitational-wave Observations and Quasi-universal Relations to Constrain the Maximum Mass of Neutron Stars,” **Result < 2.16 M**$_\odot$ – 2 separate groups and kilonova modeling. [note: pulsar PSR J0348+0432 has 2.01 solar masses]. Astro P.J. Lett.
2. “The afterglow of the LIGO/Vigro neutron-star merger (August, 2017) has continued to brighten 4x in X-rays and also in radio light. Cause unknown (maybe jet shocked cocoon around the jet). [Astrophys J Lett 2018, Ruan]
3. Regularity of high energy photon events from gamma ray burst, China: Lorentz violation as nu(E) = $c(1\text{-}E/E_{LV})$, $E_{LV} = 3.6\text{x}10^{17}$ GeV, effect 3-5σ for 25 GRB’s and photons over 40 GeV. (note: Planck energy is $1.2\text{x}10^{19}$ GeV). ArXiv 1801.08084
4. Long DNA is packed into a tight cylinder with the assist of a central axial helix scaffolding using ring shaped proteins: “Condensin II shapes a chromosome into large loops and then forms a helical scaffold for the loops to wind around. Condensin I subdivides large loops into smaller nested loops that allow for more space-efficient packing.” All side loops of DNA pass through a tiny ring, and triplets of these loops are combined by a more central small ring.

5. Researchers have found eight sites where thick deposits of ice beneath Mars' surface are exposed in faces of eroding slopes. The ice was likely deposited as snow long ago. The deposits hold clues about Mars' climate history and also may make frozen water more accessible than previously thought to future robotic or human exploration missions.
6. Advanced simulations show Black Hole jets’ streams gradually change direction in the sky, or precess, as a result of **space-time being dragged** into the rotation of the BH. Most galaxies’ central SMBH have tilted disks rotating about a different axis than the BH spin.

7. “The Tax Justice Network estimates that the planet's wealthiest individuals and corporations have stashed between **$21 and $32 trillion in offshore tax** havens, and that by eliminating this abuse, we could end world hunger and generate hundreds of millions of jobs--- Bernie Sanders.

8. Unified origin of UHE cosmic rays, neutrinos, gamma rays—all due to cosmic rays accelerated by powerful jets from supermassive black holes.
9. The current record holder for laser power is a table-top apparatus “Shanghai Superintense Ultrafast Laser Facility” at 5-10 PW (peta-watts over femtoseconds). This is very unlike the giant lasers such as the “National Ignition Facility” (10 stories high and $3.5b). A next goal is 10-100 PW at which point it should “break the QED vacuum” (separate virtual e-e+pairs into a particle avalanche storm)

10. A new type II-P supernova has remained bright for 600 days – no current model. Original mass may have been 105 suns.

11. Gravity doesn't leak into extra-large hidden dimensions based on the last LIGO N+N collision.

12. 1801.06081 In a hypothetical universe without weak interactions (a "weakless universe") stellar evolution can still proceed through strong interactions, helium and carbon can be synthesized, neutrons would be stable, and the universe remains potentially habitable. The baryon-to-photon ratio must be different for BBN to proceed adequately.
13. The first modern Briton had dark skin and blue eyes, DNA analysis of "Cheddar Man" who lived 10,000 years ago (discovered in a cave in 1903).
14. arXiv:1801.07820 [Evidence for Declination Dependence of Ultrahigh Energy Cosmic Ray Spectrum in the Northern Hemisphere. 23,854 events above $10^{9.2}$ GeV shows ankle at $10^{10.7}$ GeV for GZK effect (strong decline in proton counts due to cosmic rays on CMB photons).
15. **2 out of 5! At least 61 per cent of people who try their first cigarette become, at least** temporarily. Nicotine is highly addictive—Drug Lords.

Physics Notes January 2018

Dave Peterson, 12/12/17 – 1/12/17

1. **GW170817 binary neutron star merging** from 130 mlya gave a gravitational wave signal followed within 2 seconds by a Fermi γ-ray and then visible signal (same speed as light to 10^{-15}). This rules "out a significant fraction" of dark energy theories having a scalar field coupled to gravity" -- Some "disformal theories" and "covariant Galileon model parameters" Viewpoint aps.org
2. **Gluons** provide half of the proton's spin (lattice calculation), quark spin 30% and orbital angular momentum of quarks and gluons 20%. Original prediction was for all quarks 100%.
3. A **proton** is often pictured as quarks with gluons in a Δ configuration. Bissey (2006) found no evidence for this. The proper arrangement is a Y shape (ArXiv 060616, better than T and L shapes too). "We find a universal string tension – For large quark separations (more than 0.5 fm from system center), "the ground state potential is that which minimizes the length of the flux-tube.

4. When black holes collide with spin flip, recoil can result with enough speed to escape a galaxy 0702133. The mass of a BH is the spacetime vacuum around it.
5. 90% of Americans function in literacy and numeracy at a 6^{th} to 8^{th} grade level or less 50% below 6^{th}.
6. Our sun has a solar cycle of 11 years. There is a similar star in Cygnus 120 lya but with double the amount of heavy elements—and it's cycle is 7.4 years.

New Notes:

In 1994, the ASCA mission first detected a strong Fe Kα signal in Sgr B2, the most massive molecular cloud in the Galaxy, which is located at a projected distance of about 100 pc from Sgr A*. [Note: Kα just means 2p→ 1s—Lyman α if hydrogen]. Considering the estimated energetics of the illumination, the most plausible explanation is that the source is Sgr A⋆. Therefore, it is possible to probe the past activity of the Galactic supermassive black hole over the past few centuries by monitor- ing the echoes of its past flares while they propagate through the CMZ. The current distribution and evolution

of the 6.4 keV bright clumps indeed suggest that Sgr A⋆ experienced at least one, and probably two, powerful outbursts (L ~ 1039 erg s−1) in the past few centuries ---
From this standpoint, the molecular complex Sgr C is a highly valuable object of study
Conclusions. This work shows that Sgr A⋆ experienced at least two powerful outbursts in the past 300 years, arXiv:1712.02678

Physics Notes December, 2017

Dave Peterson, 11/17/17 – 12/12/17

Summary over last Month:

1. "h" as a constant came from Wilhelm Wien Law in 1896 [$I(v,T)=2hv^3e^{-hv/kT}/c^2$]. Planck replaced the $e^{-hv/kT}$ decreasing exponential with factor $1/[e^{-hv/kT}-1]$. And Let this = $-1/(1-x) = -\Sigma_{n=0} x^n = -Z$ (the partition function).
2. Handheld $100 Muon detector: "At sea level, you might see one count every two seconds, but on a plane at cruising altitude, that rate **increases by about a factor of 50** -- a dramatic change," http://www.cosmicwatch.lns.mit.edu/about
3. Dirac's Grave says **$i\gamma \cdot \partial \psi = m\psi$** 1984.
4. Alternative calculations for Casimir exist without ZPFs. Heuristic Casimir effect (oversimplified) uses infinite conductivity plates ($\alpha \rightarrow \infty$) but no exchange of virtual photons between plates ($\alpha \rightarrow 0$) # Nikolic agrees

5. "Gravity Sucks" might be literally true? At the event horizon of a black hole, space is falling at the speed of light. Outside v<c but inside v> c! [pg 156 Hamilton "Light emitted outward at the horizon just hangs there, barreling at the speed of light through space that is falling at the speed of light." !!]
6. The biggest tugboat for Earth's peculiar velocity (meaning on our cluster Laniakea) is the **Shapley Supercluster,** a behemoth of 50 trillion solar masses that resides about 500 Mℓya (and not too far away in the sky from the Vela Supercluster).
7. The possibility of traversable wormholes in general relativity was first demonstrated in a 1973 paper by Homer Ellis (CU).
8. In string theory, the original John Schwarz 1984 motivation of type-1 anomaly cancellation is now abandoned as just a red herring. The decisive success is really the personality of **Ed Witten** jumping into the arena.

9. Physicists were apathetic about Oppenheimer's now famous 1939 paper on gravitational collapse. It wasn't until **1957 and computer technology that MANIAC confirmed Chandrasekhar and Oppenheimer.**
10. Near PeV neutrinos seen by IceCube have cross sections enhanced by a factor of a million (they cannot pass through Earth).
11. Our Visual Horizon: observable universe radius **47 blya** (including stretching). EdS estimate is $3ct_H$ = 41 blya vs Particle horizon = ct. For our ΛCDM universe, there is a limit of 5+ Gpc~ 16- **17.3 Gℓy** for a far future **"Event Horizon"** "largest comoving distance from which light emitted now can ever reach us in the future."
12. Newtonian Gravity $\nabla \cdot g = -4\pi G\rho$, $\rho_G = -g^2/8\pi G$ leads to $E_G = -3M^2G/5a$. Large radius a implies E_G~0. Core collapse liberates gravitational energy $E_{SN}\sim GM^2/R_{NS} \sim 2.7\times10^{46}$ J $(M/Msun)^2$. A 10km NS spin energy ~ 10^{45} J.
13. New primitive Black Hole found: 690 Myr ABB 800 M suns. No one knows how such an early BH could have formed.

<u>A few Highlights from 6 years of book reading notes:</u>
Human made technosphere ~ 50kg/m^2 of Earth's surface! (2017)
QHE 2D Landau levels are Bohr orbits ($C = 2\pi r = n\lambda$)
Nuclear tests since 1945 = 2,054 (+)

Physics Notes November 2017

Dave Peterson, 10/17/17 – 11/10/17

1. A new double trap chamber has held antiprotons for 405 days enabling precision measurement of its magnetic moment as the same magnitude as that of a proton (in agreement with CPT invariance).
2. Shoji Torii (Japan equipment on the ISS) has successfully carried out the high-precision measurement of cosmic-ray electron spectrum up to **3 tera electron volts (TeV)** (Source? Pulsars? DM?).
3. Scientists have long recognized six living species of great ape aside from humans: Sumatran and Bornean orangutans, eastern and western gorillas, chimpanzees, and bonobos – now add a new but endangered Samatra "Tapanuli" orangutan species.
4. Genome sequence analysis reveals a split between early southern San peoples and other Africans at ~ **300,000 years ago!** The split between non-Africans and East Africans may be ~ 80,000 years ago.
5. SDaily: LMT (Large **millimeter** telescope – just coming online) sees galaxies back to 12.8 blya at z=6 soon after BB. The best seen spectral lines are for carbon monoxide.
6. Previous WIMP searches > 10 GeV. But He ions on the surface of superfluid He4 could detect down to 600 keV—a new realm to explore.
7. "Deep Nature Appreciation" is the "<u>feeling</u> of rapturous amazement" at the things of Nature. Human analytic reasoning is insufficient to comprehensively understand Nature so that how things work is "essentially magic" (there should be a word for that).
8. Dinosaurs would have survived if asteroid hit Earth elsewhere, scientists argues that only a few locations on Earth could create soot clouds that killed the dinosaurs.

<u>Balzac</u>: "Laws are spider webs through which the big flies pass and the little ones get caught." Marcuse: "Law and order are everywhere the law which protect the established hierarchy."

Physics Notes October 2017

Dave Peterson, 9/21/17 – 10/17/17

1. A team of scientists at the Laser Interferometer Gravitational-wave Observatory (**<u>LIGO, Virgo</u>**) announced Wednesday (9/27/17) they successfully detected another gravitational wave — the **<u>fourth</u>** so far — after two black holes collided in space. GW170814, named after the day it was detected, is a black hole collision that happened 2 billion years ago. The two black holes involved were estimated to be roughly 31 and 25 times the mass of our sun.
2. **<u>NOBEL</u>** prize in Physics: The American physicists Rainer Weiss, Kip Thorne and Barry Barish were honored for dreaming up and realizing the experiment that confirmed the existence of gravitational waves. 10/3/17.
3. **<u>10/16/17: #5 .</u>** The joint observation of GW170817 and GRB 170817A confirms the association of SGRBs with BNS mergers. Many pubs: AstrPJ.

4. Neanderthals had a greater cranial capacity than today's humans. Neanderthal adults had an intracranial volume of **1,520** cubic centimetres, while that of modern adult man is 1,195 cubic cm.
5. **NOBEL 2017 Circadian Systems:** Working with fruit flies, the scientists isolated a gene that is responsible for a protein that accumulates in the night but is degraded in the day. Misalignments in this clock may result in medical conditions and disorders, as well as the temporary disorientation of jet lag that travelers experience when crisscrossing time zones.
6. **Atmospheric carbon** 720 Gtons = 0.7Pg, Oceans 38.4 Pg! Terrestrial biosphere 2 Pg (25% above [plants], 75% below [soil]), Fossil fuels total = 4.1 Pg . CO2 released from soil in 2008 was 10x that from burning fossil fuel! (maybe from increased temperatures for rates of decomposition of soil organic matter).
7. There exists "PSA" in Planck data = large scale Power Suppression Anomaly= **weak power for ℓ <3**0 at 3σ significance. (might be relic of pre-inflationary dynamics). 1710.00759
8. Cosmo Question: Why can't neutrons in a Neutron Star decay? Answer: Because of a "rule" (PauliEP). Degeneracy pressure in a white dwarf results from free electrons not being allowed to be too close (a max density). NSs are much smaller (1/600) , so free electrons are much more restrained.
9. SCIENCEMAG (10/13/17) VLBA observes the Scutum-Centaurus on the far side of the Milky Way focusing on water maser sources at 22.2 GHz. Trigonometric parallaxes and motions determine distances.

TYSON: Fifty inches of rain in Houston. And a hurricane the width of Florida going up the center of Florida. **This is a shot across our bow.** The beginning of the end of an informed democracy.

Physics Notes September, 2017

Dave Peterson, 8/25/17— 9/19/17

1. Many astrophysicists have become convinced that rare mergers of binary neutron stars may be responsible for gamma ray flashes and the creation of gold via r-processes.
2. Gravitational lens on distant quasar challenges cosmic expansion suggesting ~73 km/s/Mpc (like Riess but different than Planck value H_o=68) .
3. Neodymium-Iron "magnetic energy product," can be on the order of 56 megagauss-oersteds (MGOe HcBr) vs common iron magnets less than 1 MGOe. The National High Magnetic Field Laboratory in Florida produces 30T to 50T.! [Galactic magnetism is typically a microgauss].
4. Full Spectrum CRAB Nebula UV=Blue, X-ray Purple. Σ 5 telescopes. https://www.sciencedaily.com/releases/2017/05/170510140756.htm
5. SD new pulsars 1.4 ms and 2.4 ms spins (radio LOFAR) Limit 0.83 ms.
6. New 100,000 sun black hole near center milky way.
7. 't Hooft on Superstring TOE: … convinced that many of the starting points and conjectures researchers have investigated up to today, are totally inappropriate, but that cannot be helped. "We are just **baboons** who have only barely arrived at the scene of science."
8. In 1967, Freeman Dyson showed that **solid matter is stabilized by quantum degeneracy** pressure rather than electrostatic repulsion; the exclusion principle is necessary for stability (J. Math Phys).
9. 98 percent of the atoms in our body are replaced yearly raises the large philosophical question, who are we if we are not our atoms and we are not our cells?

10. Who to look up to ?: **Ed Witten** has written 110 papers with greater than 110 citations ("h-index"). Below that are: Nathan Seiberg, Steven Weinberg, Leonard Susskind, Juan Maldacena, Frank Wilczek, Joseph Polchinski, and Nima Arkani-Hamed (h~61).

ATLAS On 6 July, at the European Physical Society conference in Venice, the ATLAS collaboration announced that they had found evidence for H → bb, representing an immense analysis achievement. By far the largest source of Higgs bosons is their production via gluon fusion, gg → H → bb, but this is overwhelmed by the huge background of bb events, which are produced at a rate 10 million times higher.

the Washington Post reported that the national security state had swelled into a "fourth branch" of the federal government -- with **854,000 vetted officials, 263 security** organizations, and over 3,000 intelligence units, issuing 50,000 special reports every year.
Climate Change: Trump's careless approach to policy.
"Donald Trump's saying, 'I really don't want to do this, I've got to **throw red meat** to the crazies in my base, why don't you give me some cover here,'" Scarborough said.

Physics Notes August 2017

Dave Peterson, 8/1/17 - 8/28/17

1. LHCb sees first hints of CP violation in baryons. Cronin/Fitch studies had been from mesons. But now, **Beauty baryons** like Λ_b versus anti Λ_b decay in Run 1 shows asymmetries at 3.3σ . Run 2 coming up. CP violation is not seen in the charm-quark sector. CERN COURIER.
2. Much of **Schwinger's** 1970's criticism of QCD is still quite valid—the theory remains on very tenuous ground, and is more of a parametrization than the first-principles theory it pretends to be. GUTs and strings he found outrageous not because of their theoretical failings but because he, quite rightly, found the notion of a desert between 1 TeV and the Planck scale completely unbelievable.
3. Measured value for **antihydrogen hyperfine splitting** (expressed in terms of photon frequency) is 1420±0.5 MHz, which agrees with the measured value for hydrogen to four parts in 10,000. (Nature).

4. Quark-Gluon plasma used to be exclusively from heavy-ion collisions but is now seen in simple proton-proton collisions too. Its signature is "high-multiplicity" of enhanced production of strange particles.
5. A decade-long study of a distant galaxy has uncovered the first evidence of orbital motion in a **pair of supermassive black holes** (SMBHs). This type of black hole—often weighing more than a million stars—is found at the center of many galaxies including our own, but only a handful of SMBH pairs have been observed so far. The radio galaxy 0402+379 hosts the most compact SMBH pair spatially resolved to date, with the black holes separated by only 24 light years.
6. An international team of astronomers detected four planets with masses as low as 1.7 Earth masses orbiting tau Ceti, a star about 12 light years away from Earth and visible to the naked eye. Two of the planets located in the habitable zone.
7. Using weak gravitational lensing, the huge "Dark Energy Survey" project (DES) found **26% matter** content in the universe (below Planck 33%) and also deduced w = -1.
8. North Korea Kim's push for atomic bombs and rockets started with Regan's 1983 invasion of Grenada.

9. The W boson does not "borrow" energy, it is created off the mass shell. And, Weak charge does not exist—there is no conserved quantity associated with the weak force like there is for the other two.
10. And it's very, very hard to find an elected Republican who will call this what it is: White Christian Racial Terrorism. Hateful violence is hardly new to America. But never before has a president licensed it as a political strategy or considered haters part of his political base.
11. NATURE: "Topological physics is exploding" --robustness against fluctuations in temperature or impurities.

GOP presidents from Richard Nixon to Donald Trump have been illegitimate - ascending to the highest office in the land not through small-D democratic elections - but instead through fraud and treason.

Physics Notes July 2017

Dave Peterson, 7/4/17 – 7/31/17

1. Arkani-Hamed identified the 1974 GUT hypothesis (e.g., SU(5), SO(10)) as the starting point that led HEP unsuccessfully into a wrong basin of attraction--picking a larger gauge group, then breaking it at a very high energy scale with new scalar fields.
2. Up to half of the Milky Way is made up of matter that came from distant galaxies along powerful galactic winds after having been ejected from its home during supernova explosions.
3. A new NIST Kibble electromagnetic balance for mass now calculates Planck's constant to 13 ppb when 50 ppb is required for **a new mass standard.** An alternative standard counts the number of atoms in a sphere of pure silicon.
4. **Brown Dwarfs** were only discovered in 1995, but a new study indicates that there may be 100 billion in the milky way galaxy (about half the count of stars).
5. Quantum computing may be hyped and is still poorly defined "there is not even a consensus about what exactly could make quantum computation better than classical computation." [PW Jun].
6. Exploring cosmic origins with CORE: **gravitational lensing of the CMB** – shows that the mass of the sum of neutrino types must be < 0.2 ev.
7. 58 percent of right-leaning people believe higher education ("colleges and universities) have a **negative effect** on the country." [vs 72% favorability from Democrats].
8. CRISPR can now store short movies first encoded in DNA and then placed into bacteria (and using two associated Cas proteins Cas 1 and 2).
9. The most distant star ever spotted is **9 billion light-years away** and focused to us from a huge galactic cluster. The previous farthest star observed directly was just 55 million light-years away.
10. Using CRISPR, as a U.S. first, a team of biologists has edited a human embryo's DNA.
11. New DNA analysis vs Ancient peoples: After Bronze Age Portugal and Spain (Iberia) ~5000 kya, there was minimal Steppe invasion so that a pre-Indo-European language still exists (Basque). Also, modern Lebanese derive most of their ancestry from the Canaanites of 5000 kya. [So the Bible was wrong in claiming that "an ancient war wiped them out."] And Australians are now dated back further to 65,000 years ago.

The last slide rule was manufactured, July 11, 1976

The world's population will explode from 7.5 billion to 9.7 billion by 2050; farmers will need to increase food production by 70%.
Homo naledi (South Africa) dated to ~ 300,000 ka but small brain.
The Greatest Story Too Rarely Told: America Is an Oligarchy

PHYSICS NOTES June, 2017

Dave Peterson, 6/7/17 -7/4/17

1. Chinese satellite experiment demonstrated spooky action at a distance over a record **1200 km** between two receivers in Tibet that detect opposite polarizations [Pan using new Micius satellite]. (!!!)

2. HBT: Hanbury Brown and Twiss interferometry was demonstrated for single PHONONS from an optomechanical resonator [micro-fabricated silicon nanobeam for phonon or photon resonance .. arXiv:1706.03777]

3. Guth expansion is not due to the Higgs because it is too heavy (although minor debate still exists) – a lighter particle is needed. But B+ meson decay at CERN LHCb shows that its parameter space is now excluded.

4. 65 distant galaxy clusters observed with HubbleST up to z ~ 1.8 showed major axis **galaxy rotation alignment** that has existed for 10 billion years – perhaps since birth!

5. Two Supermassive-BHs with combined mass of 1**5 G-suns** orbit at separation of 17 light years and period 24,000 years. 12 years of observation at distance 750 Mℓy to their galaxy. "Constraining the Orbit of the S**upermassive Black Hole Binary** 0402+379," The Astrophysical Journal.

6. ADM: the precise equations formulated in the original Arnowitt–Deser-Misner paper are rarely used in numerical simulations, **most or all practical approaches to numerical relativity use a "3+1 decomposition"** of spacetime into three-dimensional space and one-dimensional time that is closely related to the ADM formulation, because the ADM procedure reformulates the Einstein field equations into a constrained initial value problem that can be addressed encoded on a computer for solution.

7. propagating **surface plasmon polaritons (SPPs)** – collective boson statistics for oscillations of electrons that propagate along a metal-dielectric interface—now show interference using two entangled photons which then form separate polaritrons using SPP "launchers" using special gratings – and later convert back to photons again for HOM tests.

8. **Newton's primary goal** was uncovering what he conceived as the "true religion" of all humanity. This was his life's work more than physics. He saw himself as the true Christian – but he refuted the divinity of Jesus and the trinity.

The Women's March mobilized the single-largest demonstration in American history.
More Russians consider **Joseph Stalin** the "most outstanding person" in world history than any other leader, according to a poll released Monday. "Russia never had a proper de-Stalinization and there is little awareness" of Stalin's crimes in Russia today.
50 percent of white Americans report feeling good about Trump's presidency !!

BOB DYLAN: http://www.alternet.org/books/bob-dylans-nobel-prize-speech-how-his-songs-relate-literature
Three books that strongly influenced his songs were: **Moby Dick**, **All Quiet on the Western Front**, and **The Odyssey.** Buddy Holly and Leadbelly influenced Dylan musically.

Simple Bianchi Identities (I and II) using **differential forms.** The first Bianchi identity takes the 3-form $D\Theta = \Omega \wedge \theta$, and the 2nd BI is $D\Omega = 0$. $\Theta = d\theta + \omega \wedge \theta = D\theta$ and $\Omega^i_j = d\omega^i_j + \Sigma\, \omega^i_k \wedge \omega k_j$

Physics Notes May-June, 2017

Dave Peterson, May 10, 2017- 6/6/17

3rd LIGO detection: 49 suns black hole (31+19 at z= 0.18 distance): The new detection, which the astronomers labeled GW170104, was made on January 4 2017.

In the Ewald-Oseen light extinction, the original wave is canceled within about a mm in. Free space distance may be 2 light years.

The cornerstones of modern "quantum optics" are the Hanbury Brown-Twiss experiments and Glauber's coherence theory (alternative paths interfere and superbunching).

The LHC went to pp because "there are so many gluons in the proton that new particle production would be dominated by gluon-gluon fusion."

The dinosaurs died out because the asteroid hit Yucatan (more CO2, sulfur, gypsum)—a few minutes difference it could have been the Atlantic Ocean (just unlucky).

SDaily: Light + electrons can become superfuid at room temperature! -using light-matter particles called polaritons. Sandwich ultrathin organic film between two highly reflective mirrors and shine light. Use a microcavity between two microscope objectives.

A summary of the republican desired agenda is at: http://www.alternet.org/election-2016/50-terrible-ideas-could-become-law-if-pence-becomes-president

Dave's "Dino for Toddlers" painting to Jolie and Mike Swisher babies' room (1201 Centaur Cir. B, Lafayette,Co, 80026, Archeopterix, Triceratops).

Physics Notes May, 2017

Dave Peterson, 4/10/17- 5/9/17

Data previously collected by the MRO and Curiosity has bolstered the theory that roughly 4.3 billion years ago, Mars had enough **water to cover its entire surface** in a liquid layer about 450 feet deep.

When a photon enters glass, it causes a wave of polarization (**polaritron**) which then exits from the glass as a photon again. But, it's not just a "photon" that emerges from glass – it is also any entanglement superpositions as well !

Ting's space station **AMS** reveals more high energy positrons than expected and 5 anti-helium nuclei—but pumps need repair with a space-walk. Ting still expects dark matter particles at 1 TeV.

A new first: "Quantum test of the equiv**alence principle** for atoms in **superpositions** of internal energy eigenstates," [F=1 vs 2 hyperfine levels] using BEC clouds of rubidium atoms (1704.02296)

Data previously collected by the MRO and Curiosity has bolstered the theory that roughly 4.3 billion years ago, Mars had enough **water to cover its entire surface** in a liquid layer about 450 feet deep.

Foucault Pendulum: After 24 hours, the difference between initial and final orientations of the trace in the Earth frame is $\alpha = -2\pi\sin(\varphi\_latitude)$, which corresponds to the value given by the Gauss–Bonnet theorem. α is also called the holonomy or geometric phase of the pendulum

"Can interacting dark energy solve the Ho tension?" (Hubble Ho= 66.9 km/s/Mpc for CMB vs H = 73 for Riess local) A new interacting DE model can predict 73 from 67 with a coupling ξ near -0.3 (2σ 's) where the interaction rate is $Q=\xi H \rho_{DE}$.

Galaxy Rotation curves now are flat, but 10 by_ago they rolled over with radius as they would due to just matter without DM. So, **DM slowly accumulates** in galaxies over billions of years. (8.2 metre Very Large Telescope in Chile, n ~ 100 old galaxies). That wouldn't happen with modified gravity.

ScienceMag there has been a 40% decline in absolute mobility from 1940 to 1980 (children earning less than their parents)

Trump does not know that he does not know this or that. Rather, the dangerous thing is that **he does not know what it is to know something.** AND he has the **"Dark Triad traits"** (narcissism, Machiavellianism, and psychopathy -- having a lack of empathy, along with a grandiose sense of self-worth paired with a power-hungry drive

The Republican Party Is Sociopathic: If You Didn't Know that Already, the Health Care Bill Should Make It Clear.

Physics Notes March 2017

Dave Peterson, 2/15/17- 3/14/17

Observation of the Wigner-Huntington transition to **metallic hydrogen** at 495 GPa with reflectivity 0.91 (predicted 80 years ago but at 25 GPa SCIENCE

HIGGS interaction is a 5th force. The idea that we could predict the standard model uniquely from string theory was string propaganda from the 1980's. The idea was allowed by compactification of a Calabi-Yau manifold for N-1 SUSY. But the LHC found no SUSY so CY is in question.

NASA's Spitzer Space Telescope has revealed the first known system of seven Earth-size planets around a single dwarf star with orbits as low as 13 days.

QHE Landau levels really do have similarities to Bohr orbits—circumference = nπ This is also true in SQUID rings. A topological property is something that stays the same if you continuously change the system: Stretching it, straining it, shaving off some layers—or really **any change that doesn't cause a phase transition.**

New Claim: natural SUSY cannot be falsified unless no gluino signal is found from a pp collider at 33 TeV.

Observations of **Ten-photon Entanglement** using thin BiB3O6 crystals ("BIBO") GHZ state 10 H's + 10 V's. 1609.00228

The wave mechanics of α –ray tracks," 1929 N.F. Mott is now considered the first pioneering example of decoherence theory where a spherical wave is converted into a ray. A vapor atom changing state or being ionized localizes the wave function. [Note: "Photons that never end" may have had such minor interactions in the past—collapse doesn't mean dead].

SD 3/1/17: 5 black hole **gamma ray blazars** have been found < 2 Gy after BB. Fermi catalogued 1.4 million quasars. A mystery is how they formed so quickly

Physics Notes January 2017

The planet's **technosphere** (human made) now weighs some 30 trillion tons -- a mass of more than 50 kilos for every square meter of the Earth's surface

Current Scientific Error: "Consciousness is the new face of vitalism."

More than 30 species of non-Avian dinosaurs have been confirmed to have feathers.

Na**mbu-Goto** equations can give Veneziano amplitudes. It took a while to quantize the Nambu-Goto action, which is possible in the "light cone gauge."

Arkani-Hamid said "Gauge symmetry is a complete fiction." (from twistor theory)

The USA has killed around 20 million people since WWII (mainly civilians).

Physics Notes December, 2016

Chomsky: Trump's win puts government in the hands of the most dangerous organization in world history. (and he means it). [Note that fewer than 25% of the people in our country voted for Trump].
Casimir force does not originate from vacuum energy but is rather a non-vacuum van der Waals force.

MUST READS: Noam Chomsky, "Trump in the White House," see Chomskyinfo.org, and see "Neoliberalism" in Wikipedia -- a policy model of social studies and economics that transfers control of economic factors to the private sector from the public sector. On Global warming: "40% of the US population does not see why it is a problem since Christ is returning in a few decades." "New Democrats are pretty much what used to be called 'moderate Republicans.' " Friendly fascism "but that requires an honest ideologue, a Hitler type, not someone whose only detectable ideology is Me."
[vs neoconservative: relating to or denoting a return to a modified form of a traditional viewpoint, in particular a political ideology characterized by an emphasis on free-market capitalism and an interventionist foreign policy.]

Physics Notes November 2016

Dave Peterson 10/26/16 -11/16/16

1. "Will astronauts traveling to Mars remember much of it?," Mars-bound astronauts face chronic dementia risk from galactic cosmic ray exposure (clearly demonstrated with tests on rodents).
2. Israel currently has the most right-wing government in its history, and "leftist" is a bona fide bad word whose definition just keeps broadening. Trump's election will solidify this.
3. Patient Zero for HIV is a myth. The American outbreak occurred a decade earlier in New York City due to a jump from the Caribbean.
4. **Anti-protons** placed into normal He+ atoms have transitions indicating exactly the same mass as protons. [2E+09 pbarHe+ atoms, sharp spectral lines to 10 ppb, 1.5 K, vs QED calculations to order 7].

Physics Notes October 2016

Dave Peterson October 1, 2016 – 10/25/16

1. A "Scientific Reports" article 10/21/16 showed **only "Marginal evidence for cosmic acceleration** from Type Ia supernovae," <~ 3σ ‼ A ten times larger data base (n~740 sn) shows results that "are still quite consistent with a constant rate of expansion." R=ct , ‼ see http://www.nature.com/articles/srep35596

2. Ed Witten (as a history major) was told to read Jackson E&M. He did so in two weeks and understood it. (hard graduate text).
3. The observable universe contains at least **two trillion galaxies**, ten times more than previously thought from pictures of the older "Deep Field" views (100 billion).
4. For the first time, astronomers have clearly observed at infrared wavelengths what happens after a black hole eats a star: it burps back up a brilliant flare of light that echoes through space (Tidal disruption Flares, NASA).
5. 7Li calculated abundance is significantly higher than the one deduced from spectroscopic observations. http://arxiv.org/pdf/1609.06048.pdf
6. There are now several stars that mysteriously disappear --might be the first confirmed case of a failed supernova, a star that tried to explode but couldn't finish the job. A newborn black hole appears to have been left behind to snack on the star's remains.

7. We Just Passed A Grim Carbon Dioxide Threshold, Possibly For Good
   CO_2 levels surpassed **400 ppm** in September. Scientists say we won't see a month below that symbolic benchmark "ever again."
8. Charge Radius Experiment with Muonic Atoms collaboration has measured the radius of the deuteron more accurately than ever before, finding that it is significantly smaller than previously thought. And of course the muonic deduced proton size is smaller than the electron deduced size. This is a major mystery! CERN.
9. David Thouless, Duncan Haldane, and Michael Kosterlitz won this year's **Nobel** Prize in Physics for their theoretical discoveries using topological concepts. Their work pioneered a new understanding of phase transitions of matter,
10. The NSA has access to virtually every phone conversation in the US. And half of us are now in the facial recognition system.
11. Putin and the oligarchs who surround him often argue that the United States is **"hypocritical" in advocating democracy, transparency, and human rights,** because our own practices are imperfect. And Iran's Hassan Rouhani said that "morality" doesn't exist in the U.S.
12. SCIENCE insists on truth, honesty, integrity, persistent analysis, non-authoritarian – but Trump/Republicans are authoritarian, lacking in integrity, highly biased (e.g., Racist), not very analytical -- Hence <u>intrinsically anti-science</u>.

<u>Physics Notes September 2016</u>

Dave Peterson, 9/2/16 – 9/30/16

1. Sabine H Back Reaction: The 2016 **750 GeV** bump has declined. The LHC nightmare scenario has come true. **<u>The Higgs and Nothing Else!</u>**
2. There is a <u>dim nearby galaxy</u> that is nearly 100% dark matter—Dragonfly 44 in Coma Cluster < 1% of the stars in the MW but nearly all the mass.
3. The planet is **warming** at a pace not experienced within the past 1,000 years, at least, making it "very unlikely" that the world will stay within a crucial temperature limit agreed by nations just last year. And variations are clearly larger too: see visual data presentation in: https://www.theguardian.com/environment/2016/aug/30/nasa-climate-change-warning-earth-temperature-warming
4. **Ramanujan** said that a Hindu Goddess Namagiri appeared in his dreams, whispered equations into his ear and showed him visions of scrolls covered with strange formulas.
5. Heisenberg's visit to Bohr in 1941 resulted in misunderstanding: atomic bomb cost and difficulty for Germany made it "impracticable while the war lasted". So the Nazi Government formally dropped it in 1942 leaving just minor reactor work. Heisenberg was contemplating morality of potential development much later on after the war, but Bohr didn't know this. Then, Manhattan project physicists believed they <u>were "in a race with Heisenberg" that wasn't real.</u>

6. Cooper pair bonding uses a "significant delay from ion vibration coupling" (slow phonon) seemingly entanglement action at a substantial distance (weakly bound state meV < 1 micron separations). The pairs act as bosons. Recently, large Rydberg atoms bond over microns distance (huge).

7. The Kuwait to Iraq "Highway of Death" mass murder was declared a war crime in violation of safe passage protection under UN Resolution 660.
8. New reduction of Alzheimer plaques accomplished by "antibody aducanumab" (Nature) and also separately by commonly used NSAID mefenamic acid. But, is removal the same as "cure?"
9. Newly revealed most distant galaxy cluster CL J1001 at **11.1 b ly** shows very high star production rate.

Physics Notes August 2016

Dave Peterson, 8/1/16 - 9/2/16

1. **Violation of the Leggett-Garg Inequality in Neutrino Oscillations,** PRL 7/26/16: MINOS 6-sigma result over 735 km record distance! LG involves correlations of measurements on a system at different times—in this case an ensemble of neutrinos. The coherence length of neutrino oscillations can be over vast distances.
2. Arxiv **Super-Kamiokande-IV** sees low energy 8-B solar neutrinos down to 3.49 MeV and gets a measured solar neutrino flux 2.3E+06/cm^{2}sec, day/night asymmetry -3.6% (± ~ 2)
3. SN: Scientists find clue to why mitochondrial DNA comes only from mom Scientists have identified a protein that chops up the mitochondrial DNA in a dad's sperm after it fertilizes an egg. The finding helps explain why mitochondrial DNA is usually passed on only by mothers.
4. 1607.02240 Khakimov ANU **Ghost imaging with He BEC** collision of 2 beams – one through ANU mask to a bucket detector to a correlator.
5. 7/11/16: the **750 GeV** bump is vanishing at CERN! 500 papers written about a statistical fluctuation. Physicists in grief. Higgs only, nothing else!
6. There are four new names for **chemical elements:** Nh 113 Nihonium, muscovium Mc 115, tenessine Ts 117, and 118 Og after Russian physicist Yuri Oganessian, who contributed to the discovery of several superheavy elements. So Row 7 from element 103 on now reads: Lr, Rf, Db, Sg, Bh, Hs, Mt, Ds, Rg, Cn, Nh, Fl, Mc, Lv, Ts, and Og 118.
7. TIME: In the study, Khalili studied **CRISPR's** ability to remove HIV from both mice and rat models, and found that overall, it was successful in cutting out the virus in more than 50% of the cells of each type. Surprisingly, he achieved this with two simple injections of the molecular CRISPR scissors into a vein in the animals' tails.
8. **Onion**: a) Nation Surprised It Took So Long For Primaries To Weed Out Candidate With Genuine Principles. B) Trump: "It's completely shameful to take words I've spoken or written and try to connect them to some kind of objective reality. I say something, and the next thing I know, a crooked reporter is telling everyone what I said along with a fact-based explanation of what its implications are and why it matters. It's ridiculous, and it has to stop." C) Rep Convention: Bill Maher: "Did you see Donald Trump's speech? If that speech was any darker it would've been shot by the police."

Physics Notes June 2016

David Peterson, June 20 - 8/2/2016.

1. There are four new names for **chemical elements:** Nh 113 Nihonium, muscovium Mc 115, tenessine Ts 117, and 118 Og after Russian physicist Yuri

Oganessian, who contributed to the discovery of several superheavy elements. So Row 7 from element 103 on now reads: Lr, Rf, Db, Sg, Bh, Hs, Mt, Ds, Rg, Cn, Nh, Fl, Mc, Lv, Ts, and Og 118.

New Notes:

a. The Second Amendment was ratified to preserve slavery and guarantee state slave patrol militias.
b. 7/11/16: the 750 GeV bump is vanishing at CERN.

Die Grenzen meiner Sprache sind die Grenzen meiner Welt. The limits of my language are the limits of my world. (find a word for "natural miracle," wonder, marvelous, "Nature")

Physics Notes May, 2016

Dave Peterson, 5/3/16- 6/2/16

1. The local Hubble constant of **H** ~ 73 km/s/Mpc is well above 67 from Planck (1604.01424) suggesting that late times had +8% faster expansion than expected from early times.
2. NOvA detector saw 6 electron neutrinos from a muon neutrino beam over 810 km (+ 3.3σ)
3. Conscription: There was no draft since 1973, Carter 1980 said men ages 18-25 must still register. New bills desire registering women as well as men and perhaps ages 18-73!
4. The new worth of planet Earth is over $240 trillion of which $12 trillion was looted from poorer countries who also owe $8 trillion in foreign debt.
5. Origin of Ashkenazi Jews: Eastern Turkey 700 ad in 4 primeval villages: Iskenaz, Eskenaz, Ashanaz, and Ashkuz near the silk road trade between Germany and China. Yiddish was a secret language of merchant monopoly. This origin was lost, and much later, Ashkenaz was used to refer to Germany.
6. Fact checking political truth: **True or mostly true** Sanders 54% of the time, Hillary 51%, Obama 48%, Carson 4%, **Trump 7% ‼** "The Donald" defeated his more truthful opponents (Jeb 48%, Christie 41%, Rubio 38%, Cruz 22%) ‼ Trump says whatever gets him the result he wants. He understands humans as 90-percent irrational and emotional and acts accordingly.
7. HAWC high altitude water Cherenkov gamma ray observatory Mexico 300 water tanks see gamma rays at energies of 100 GeV to 100 TeV ‼ And Ice Cube has been searching unsuccessfully for sterile neutrinos at an energy range of 320 GeV to 20 TeV.
8. Photons with **half-integer** $\hbar/2$ angular momentum have recently been produced; and the corresponding picture for "half-twists" is a double helix. "For light as well as electrons, reduced dimensionality allows new forms of quantization." http://advances.sciencemag.org/content/2/4/e1501748
9. Proof that Casimir force does not originate from vacuum energy, Hrvoje Nikolic, 1605.04143, $F(y) = -\partial E_{vac}/\partial y$, but H_{em} has no explicit dependence on y nor matter fields. This study is a separate validation of Jaffe from van der Waals forces.

Still There: http://fqxi.org/community/forum/topic/2351 "Physics Lives in Form Heaven."

Physics Notes April 2016

Dave Peterson, 4/1/16 – 4/12/16

1. **LIGO** unexpectedly saw black holes in the middle range of 10-100 suns; and interestingly, dark matter as primordial black holes in this narrow range is still allowed (not yet forbidden).

2. Clocking the rotation rate of a supermassive BH (quasar OJ287, 3.5 Glya, 18 Billion suns) indicates that a close smaller BH precesses **39 degrees per orbit!** KERR Parameter = 0.31 < 1.0.
3. "V404 Cygni, about 7,800 ly from Earth, was the first definitive BH to be identified in our Galaxy and can appear extremely bright when it is actively devouring material." Its last eruption was in 1989 and again in June 2015 revealing a quick bright red flash.
4. Legalization: Kids find it much easier to get hold of illegal drugs than legal drugs. Portugal decriminalized all drugs — and injecting drug use fell by 50 percent. Switzerland legalized heroin for addicts over a decade ago. Nobody has ever died on an overdose there on legal heroin. A Harvard Professor calculates the murder rate would fall by at least 25 percent after legalization.
5. Fraassen: there are (at least) two levels of reality: One consists of the rules and regularities of the physical world, which science can access and measure. But the other level, the ultimate source of those rules and regulations, science can never even access, much less come to know. - See more at: http://www.space.com/32452-can-science-explain-the-multiverse.html#sthash.tyDY1gGS.dpuf Also, David Gross hates the anthropic principle.
6. In hopes of understanding sphalerons, I spent most of this month studying background math: topology, homotopy, homology, fiber bundles, invariance classes (e.g., Chern classes, Chern-Simons forms).
7. "Disaster on Earth" perhaps 70,000 years ago the giant Sumatran **Toba** supervolcano seems to be responsible for reducing humanity to below 10,000 surviving individuals (supported by genetic evidence) – Toba bottleneck theory, ash and cooling.
8. Retrocausality reading: http://www.ijqf.org/forums/forum/quantum-foundations-workshop-2015/retrocausality-and-transactions Also O. Costa de Beauregard's Parisian Zig-Zags, ~ 1950.
9. Option key for Word symbols: ÷ = option /, åÅ = aA, ∂ = d, ß = s, ü =opt u u, ö = opt u o,
10. J. J. Thomson won the 1906 Physics Nobel Prize and his son George Thomson won the 1937 prize: "the father got the prize for showing the electron is a particle; the son for showing it is a wave".
11. PANAMA: http://panamapapers.sueddeutsche.de/articles/56febff0a1bb8d3c3495adf4/
12. The strongly sexually repressed state of Utah has the highest rate of online porn web subscriptions.

Summary for March, 2016:

1. **Extinctions** and Volcanoes: 5 large volcanic explosions match extinctions very well: **Siberian Traps** LIP (large igneous provinces) at end of Permian 252 Mya, Viluy traps (late Devonian 373 Mya), Central Atlantic LIP (end of Triassic 201 Mya—eg now Florida) and **Deccan Traps** LIP India (End of Cretaceous 66 Mya!). [in addition to giant impacts!].
2. Wanted! More Bumps: there is a bump in prompt energy antineutrino data from Daya Bay at **5 MeV** (positron spectrum) that has been seen before but is now at 4.1 σ (before at Double Chooz and at RENO).
3. Fermilab bags a **tetraquark** (udsb), the first with all different flavors X(5568) 5.1σ and without charm or antiquarks!
4. **High energy cosmic gamma rays** are now routinely seen above 100 GeV (~EWSB transition). The Crab nebula is a well-known emitter of gamma rays above 1 TeV (first detected in 1989). Now there are over a thousand known **TeV**

gamma ray emitters. There are clear gamma ray energies above 10 TeV, but the case for > 100 TeV is not yet established.

5. Record distance to a galaxy: EGSY8p7 had z = 8.68 (13.2 Gya), but the new GN-z11 has z = **11.1**! (13.4 Gy ago)!
6. For our ΛCDM universe, there is a limit of **17.3 Gℓy** for a far future "Event Horizon" (limit of possible observation distance forever) which also applies to the future Hubble distance as well.

Summary from February, 2016:

1. "**Observations of Gravitational Waves!** from a Binary Black Hole Merger," PRL 116, 061102 (2016) from event on 9/14/15. (now want n = 2)
2. There now seems to be a new intermediate 100 K suns **black hole** only 200 ly from Milky Way center Sgr A* (which itself has a mass of 4.3 Million suns).
3. Iran removed the core of its plutonium Arak reactor and filled it with concrete rendering it harmless and paving the way for economic and financial sanctions to be lifted soon.
4. Gordon Kane 1601.07511 compactifies M-Theory on manifolds of **G2** holonomy to describe our vacuum and claims that LHC run 2 should be able to see gluinos, LSP, winos and binos. The discovery of the Higgs boson is evidence for supersymmetry, and one can calculate accurately the ratio of Higgs mass to Z mass and get 126.4
5. The top most cited papers in physics strongly include the arena of **DFT** (density functional theory): Becke, Kohn, Perdew. (e.g., Kohn-Sham was used in more than 30,000 papers in 2015)!
6. The highest redshift of SNe 1a is z =1.914. Cosmic Coincidence: the average acceleration of the universe up to the current time was near zero!
7. Quotes: "Fine tuning requires no special explanation at all, since it is not the Universe that is fine-tuned for life, but life that has been fine-tuned to the Universe." (Klas Landsman). "Inflation isn't falsifiable, it's falsified!" (Penrose, 2015). "BICEP did a wonderful service by bringing all the Inflation-ists out of their shell, and giving them a black eye."
8. Other Short notes: "Trump is a bonfire in a field of damp kindling" (and wins by sidestepping middlemen). His campaign is instinctual stream-of-consciousness. He is a narcissist (a personality disorder, like LBJ) – **but** half of previous presidents also had major psychological problems. The richest human was Mansa Musa of Mali (~$400 billion, 1300 ad.) Hitler admired America's KKK, Jim Crow, and Indian genocides (so maybe, in part, evil is being a century behind). Artificial intelligence machines can now beat humans at **Go**.

Summary from January, 2016:

1. We can now see inwards towards the M87 giga-black hole to 5.5 Schwarzschild radii. (Goal: see closer in).
2. Elon Musk's Space-X achieved its first precision historic landing 12/22/15 by using grid-fins that fold out after separation.
3. Optical focusing is achieved when its ∫ intensity squared is at a maximum.
4. Conservatives have a personality "negativity bias" tuned to threats and disgust and gave survival value during the Pleistocene. It is a different way to perceive the world and is relatively immune to reasoning.
5. The genome editing method **CRISPR** is Science Magazine's 2015 breakthrough of the year and cut costs a thousand fold. First found and applied in Yogurt bacteria, it is now used in mouse gene editing (e.g., muscular dystrophy genetics) and making Mammoth DNA from elephants.

6. Maxwell's 1865 field equations used **20** field variables with names like PQR, pqr, FGH, fgh, α β γ and was re-written in 1873 for quaternion form. Heaviside/Gibbs altered this to vector analysis in 1884.
7. There is no currently accepted potential shape for inflation, and there is still a big problem with fine-tuning and Planck2015 results giving tensor/scalar $r < 0.1$ when gravitational waves should be much stronger than that from magnified early quantum fluctuations (that also produce variation in CMB temperatures).
8. New chemical elements have bee identified but are still unnamed by IUPAC: Z = 113 (RIKEN) and 115, 117, 118 (DUBNA)

Summary from December, 2015:

1. LHC CMS and also ATLAS both show new boson bumps at **750 GeV** for di-photon plots (see Strassler, Carroll). Hundreds of articles also now appear about the possible nature of the 750 GeV bump (…ambulance chasers?).
2. W': The LHC has finished its 2015 run colliding protons at 13 TeV, will now turn to heavy ion physics. There are ~ 2 TeV bumps suggesting new particles, and the (1510.08083) W' goes with **SU(2)R!** Extension of the SM (as in LRS = Left-Right Symmetry). Not yet confirmed and not clear that this bump will survive further stats!
3. LAGEOS (and –S2) satellites convincingly show gravitational rotation frame drag (2 meters net over 11 years, Ciufolini).
4. Physics World 2015 Breakthrough prize is for Double-Quantum-Teleportation by Pan and Lu (both spin and OAM). Best Book award to **Amanda Gefter!**
5. There have been **2,054 nuclear tests** since 1945 — many near populated places.
6. Obama the most threatened President in history: Since the President took office in 2008, the rate of threats against the president has increased 400% cent…. That's 43,830 death threats for his first four years alone.
7. **Law of Jante**: is the prevalent idea that there is a pattern of group behaviour towards individuals within Scandinavian communities that negatively portrays and criticises individual success and achievement as unworthy and inappropriate-- a mentality that de-emphasises individual effort and places all emphasis on the collective, while discouraging those who stand out as achievers.
8. Rename Homo sapiens ("wise"—our least evident feature) to Pan narrans (storytelling chimpanzee).
9. Princeton advanced study: the steady state of a mathematician is to be blocked.
10. See fascinating WIK Blue Dragon Images—Glaucus Atlanticus.
11. The sinking of a Spanish multi-billion dollar treasure ship (5/28/1708) by the British near Cartagena drastically defunded the war from France and Spain, altered the course of history and led to the American revolution.

Summary from November, 2015:

1. A black hole event horizon is a view from infinity (far away). If one continually calculates this absolute horizon for in-falling matter or in-falling shells of matter, the **horizon expands outwardly in anticipation** and does so quickly enough to ensure that no matter ever needs to penetrate it. [Matter doesn't cross the horizon; the horizon expands to engulf the matter].
2. Arkani-Hamed is pushing a 100 TeV "Great Collider" for China 60 miles in circumference.
3. Physicists were apathetic about Oppenheimer's now famous 1939 paper on gravitational collapse. It wasn't until 1957 and computer technology that

MANIAC confirmed Chandrasekhar and Oppenheimer. [Claim Blandford: GR around a BH can use a Newtonian type potential V = -mG/(r-2m)].

4. CURES: Schizophrenics can be taught early to tolerate the voices in their head (that may always be there) ..Program "RAISE" . Drug Vivitrol can safely cure alcoholism. Vitamin C can indeed cure a few types of colon cancer. Oxytocin can help Autism.
5. The Higgs field supplies weak charge to the vacuum. Fermions gain mass by zig-zags grabbing weak hypercharge (Y) from the vacuum and pushing out weak isotopic spin (I3) so that Y can go up when I3 goes down and vice versa (**both** of these types of charges). This **sea-sawing** is required to conserve electric charge Q = Y/2 + I3.

Summary from October, 2015

1. Opera Bags **Fifth Tau Neutrino** (N = 5 and also 5σ confidence also). Data 2000-2012 Gran Sasso 730 km travel with muon neutrinos converting into **tau** neutrinos during travel from CERN Aps.org Phys. Rev. Lett. **115**, 121802 (2015)
2. Finally an answer on **covalent bonding:** Frank Rioux . Delocalization & Born enhancement. http://www.users.csbsju.edu/~frioux/h2-virial/virial-h2.htm The covalent bond clarified through the use of the Virial Theorem (from Slater).
3. **Police only shot four people** (and nobody died) in all of England the past two years.
4. Possible Dark Matter **decay** Signal: We detect a line at 3.539±0.011 keV in the deep exposure dataset of the [Milky Way] Galactic Center region, observed with the XMM-Newton. (also Andromeda and Perseus cluster)
5. Only about a third of college graduates (during the past 10 years) agree that their education was worth the cost.
6. **Nobel Prize Physics:** Japanese scientist Takaaki Kajita and Canadian scientist Arthur B. McDonald for their discovery of neutrino oscillations, "which shows that neutrinos have mass." [from Super-Kamiokande and SNO Sudbury]. Peace prize to Tunisian groups.
7. Jewish physicist Bruno Touschek [of electron-positron collider fame] was employed in Germany in 1944 to construct a Wideroe type **betatron**. When finished he was arrested and shot in the head and left by the side of a road to die —he survived and was later employed by Amaldi in 1952.

Summary from September, 2015:

1. Strong Λ_QCD value scales with energy too and with the number of quarks below it (low energy n = 3 quarks, Λ ~ 350 GeV, but at Z scale 5 quarks Λ ~ 217 GeV).
2. I sent a Challenge to Dr. Ruth Kastner: 1) about majority cosmic photons that "never end," 2) Born rule causing covalent bonding without obvious transactions. But I also learned a new rule: electron delocalization creates more volume in which to move and this lowers atomic KE.
3. Rovelli/Lao Tzu: "The wave function that can be told of, is not the True Wave Function."
4. Question from Cosmology: If an electron is a wave, how does it spin? (Well, being a particle doesn't really help either. And a photon has a rotating A field giving rise to E and B squared energy fields that rotate).
5. **CNB** = Cosmic Neutrino Background = (4/11)^(1/3) of the cosmic microwave background (CMB) temperature **~1.95 K** for the CNB. Joint interactions ceased at 1 second after Big Bang. It might be possible to verify this by peak shifts of the CMB from neutrino effects.

6. 2-slit Bohmian trajectories cannot cross each other—but that makes little sense in a beam splitter (bounce versus transmit).
7. There has been a major shift in potential climate future: China has gotten on board with less coal use and efficient energy use development such as LEDs—Less Hopeless.
8. 501.05658 Avoiding **Haag's Theorem** with Parameterized QFT. "any field unitarily equivalent to a free field must itself be a free field." "Haag's theorem is very inconvenient; it means that the interaction picture exists only if there is no interaction." Axiom changes: Uniqueness of the Vacuum |0⟩ requires different |0⟩_interaction, and evolution is now off-shell.

Summary from August, 2015:

1. **Pentaquark**-charmonium states have been seen with 9 σ confidence by LHCb near Pc 4.4 GeV for Λb baryons decaying into J/ψ's, protons and charged K.
2. Ionized Buckminsterfullerene (C60+) molecules have been seen in the gas of space (astronomical versus lab measurements).
3. IQ estimates: Einstein ~ 160, Galileo ~ 182, Newton ~ 190 (and mathematician Terry Tao ~ 230 -- at age 8 he had math SAT score of 760).
4. Australasian ancestry seen in some Amazonian tribes and also contemporary Aleutian islanders (perhaps 23 kya to **Beringia** then 15 kya to Americas).
5. It was dogma that constrictor snakes kill by suffocation but it is now known that immense pressure causes cardiac arrest instead.
6. From a Nun: pro-life should not only be child born but also child fed, child educated and child housed as well.
7. Reference: StdModel 100 pages 0001283, quantum cryptography 1508.00341,
8. MicroBooNE sees first cosmic muons in 170 tons of liquid Argon. Primary solar pp neutrinos seen at Boraxino.
9. Wilczek seems to now have replaced the term "GRID" with ``Property Space''.

Summary from July:

1. Sabine Hossenfelder's Blog Back Reaction on quantum gravity: the fundamental problem is that we haven't understood quantization or the quantization prescription.
2. **Pluto** can now claim to be king of the Kuiper belt, the region of thousands of icy worlds that orbit the sun beyond Neptune (2370 km Dia > Eris at 2326 km).
3. Five microscopic streaks over 5 years represent the conversion from muon neutrinos to **Tau neutrinos** at Gran Sasso (Opera).
4. **Neal Turok** says that the multiverse is ``the least predictive theory ever.'' After seeing no SM extensions, theorists are walking around in a bit of stunned silence.'' This is a kind of catastrophe – we've lost our way…
5. Polchinski says that monogamy breaking for Hawking radiation is like when chemical bond are broken in chemistry (and chemical bonds are indeed entanglements).
6. Special ArXiv: 1506.04120 2013 Nobel colloquium on Higgs,
7. General: a) The primary initial motivation for Escher's drawings was Moorish art such as the Alhambra. B) in 1860, US slaves were worth more than all the manufacturing companies and railroads in the nation.

Summary of June:

1. Some supernova explode asymmetrically (e.g., as seen from chemical signature of inner shell ejecta such as Titanium-44 (sciencemag). **1987a** is one such example!

2. EBL (SciAm June 2015: All the Light there ever was). Photons can collide with photons, HE gamma on LE photons. Blazars can give off GeV gamma rays and sometimes up to 20 TeV. Initial energy can be modeled (wrt rest of spectrum). 9 TeV Blazars revealed EBL collision loss.
3. Latest on Higgs: definitely spin zero (99% confidence level) and mass 125.09 ± 0.24 GeV (0.25%)
4. Most SN1a explosions come from doubly degenerate stars (white dwarfs), but some are single-degenerate (WD+sun like with bright UV).
5. Briefs: 42% of new grads will make less than $25K. A warm-blooded fish has been found (Opah=Moonfish). Kurchatov died from radiation poisoning from the Chelyabinsk-40 catastrophe of 1949. The John Templeton Foundation has $3.34 billion dollars!

Summary from May, 2015:

6. The **SN-1987a** burst resulted in 24 detected antineutrinos within 13 seconds interval on Earth. Sensitivities were altered by how much earth they had to pass through (MSW electron density effect). This is analogous to having an index of refraction for neutrinos (typical n-1 ~ 10^-20).
7. Electron beams can become twisted by passing over a monopole (e.g., tip of a magnetized needle), and these vortex beams can detect chiral crystal handedness in thin samples.
8. Cosmic rays on Mars penetrate 1 meter of surface and would kill life! There may be the equivalent of 1 m of ionic-water-ice covering Mars (with calcium perchlorate).
9. There are two types of SN1a explosions with different (standard) candles. This forces a correction to dark energy content of the universe.
10. Penrose believes in the quantum information time-reversible zig-zags as in EPR.

11. Bacteria (immune system) can remember the DNA of viruses that have attacked them and can then slice the bad DNA. This idea can be applied to editing mammal DNA too (called **CRISPR**).
12. He-4 atoms launched into the two input ports of a beam-splitter always bunch together in the same output port (Hong-Ou-Mandel effect) Nature Letter.
13. The inflaton decays, but we only see the end effect because prior decay is highly diluted by expansion. Current curvature radius is > 4 x radius of our observable universe (from CMB data, see Physics Today 3/15 p28).
14. The Virgo Cluster of 2000 galaxies spans 10 times the angle of the full moon in our sky (only 70 Mℓy distant).

Summary from April, 2015

1. Enormous water loss over Martian history can be seen in the ratio HDO for Mars/Earth ~ 7. The H2O went out into space.
2. The human-dominated geological epoch known as the ***Anthropocene*** probably began around the year 1610, with an unusual drop in atmospheric carbon dioxide and the irreversible exchange of species between the New and Old Worlds, according to new research. From Christopher Columbus to 1610, 50 million Native Americans were killed resulting in trees taking over their land. This in turn sequestered more carbon and caused a global cooling.
3. BICEP2's data conclusion showing unexpectedly strong B-mode signal was postponed for a year for re-checking. They were blocked from seeing the needed dust data from ESA's Planck satellite (which would have saved embarrassment.

4. Nicholas Gisin (Quantum Chance, book on Bell entanglement) says ``that there is no explanation in the form of a story taking place in space as time goes by'' – so non-local correlations are outside of spacetime!
5. Kennewick Man (9000 years ago in America, bones found 1996) was finally studied (lawsuit by anthropologists against the Army corps of engineers and Justice Department) and found to be unrelated to any living indian tribe but most closely resembling the Ainu of Japan.
6. Zoltan Fodor's giant LQCD supercomputer calculations finally showed why the mass difference n-p is 0.14% (proton charge contributes +0.1% but d-quark > u-quark).

Summary from March, 2015:

1. Neutrons are now being used for Bell Tests (1502.07338) 6.25A neutron beam on Si scatterers and polarizers – shows effect, but more development needed.
2. Poem from early Schrodinger class: ``Erwin with his psi can do, Calculations quite a few. But one thing has not been seen, Just what does ψ really mean?''
3. Active-SETI is now being discussed—broadcasting out to potential listeners. As a joke, suggested messages include: Klaatu Barada Nikto, and ``Is your quantum state ontic or epistemic?"
4. EARLY AGES: New squirrel like mammals now date back to 210 Mya (dinosaurs 230 Mya). Ancient rocks now suggest nitrogen processing – life on earth at 3.2 Gya! Epoch of reionization started 550 My after BB. The ratio Mp/me was still constant within 1ppm 12.4 Gya (e.g, quasar z=4.42 shining through a galaxy with z=4.22).

5. The Lamb Shift is real and does contribute to gravitational mass (as shown in Al and Platinum M_g/M_i despite vacuum polarization differing by a factor of 3).
6. Humans and proto-humans depend on high calorie sources for their brains. They deleted two bitter taste genes (note that wild yams and tubers are bitter but taste better after cooking).
7. Kastner: internal Feynman lines do not prompt an absorber response and are time symmetric. Haag's theorem encourages direct-action-theories – a Fock space description of interactions doesn't exist. (1502.03814).
8. After 8 centuries, rats exonerated in spread of Black Death. Gerbils implicated.
9. WOIT: **AdS/CFT has 10K papers!,** but toy model case AdS3/CFT2 shows how little seems to be truly understood. I think this is an historically unprecedented situation.
10. 49% of Republicans say they do not believe in evolution. Only 37 percent say they do. 66% say they do not believe in global warming. 57% would support establishing Christianity as our "national religion."

Summary from February, 2015:

1. Eddington characterized the Einstein static universe as matter without motion and the de Sitter world as motion without matter. There are different forms of the dS metric (static vs exponentially expanding). Let the form of g11^-1 = (1-r^2/α^2) = (1-Λr^2) = (1-H^2R^2). The original static closed dS space only had g11 ≠ 1 (no time dependence) . The exponential expanding space has this and goo = 1/g11. See: http://sigmapisigma.org/radiations/2008/ecp_bigbang3.pdf

2. 't Hooft doesn't believe in Hawking radiation (it is just semi-classical but needs full QG) and also believes that there should be a deterministic theory underlying QM.

3. Encroma now makes glasses to correct for RG color blindness. (Trick, precise spectral cuts, block UV, Cyan, and Yellow out to leave RGB).
4. SUGRA is the "Dirac square root" of GR (1501.03522). [Deser, ADM].
5. There are now 89 galaxies with reliably measured mega black hole masses (1501.02937). Astronomers have now found over 1,800 exoplanets (SciAm 1/15).
6. Sabine H: string theorists don't actually do string theory any more, they do AdS/CFT (with over 10,000 citations).
7. Advice to Russian travelers in America: ``Did you know bribery was illegal in America? Be careful of that." And you must keep them smiling—don't say anything negative.
8. Smoot also added that positive and negative energy summed to zero in our universe (still confusing). But, see Joseph Silk, The Big Bang, 2001 p 95, In a Newtonian expanding cosmology, the sum of KE and negative PE doesn't change, conserved.

Summary from January, 2015:

1. One of mankind's most ancient continuous lineages is the African Khoisan tribespeople (Bushman) distinct from Europeans, Asians and all other Africans and going back at least 150,000 years. (ScienceDaily 12/4/14 DNA shows no admixing).
2. ScienceMag: Full DNA on 48 bird genomes reveals bird history with the "least radiated clades being Chicken, Turkey, Ostrich and Tinamou. Birds originated from a theropod dinosaur lineage more than 150 million years ago during the Jurassic and are the only extant descendants of the dinosaurs. The common ancestor lacked mineralized teeth. Birds possess the most advanced vertebrate visual system with more color types than mammals.
3. The death of the big dinosaurs was due not only to the Chicxulub asteroid impact but also to the Deccan volcanic eruptions lasting nearly 750 kyr during the same time (66 Mya).
4. SciAm Jan 2015: Earth has borderline capability for intelligent life. Planets with 2 earth masses and a smaller weaker reddish K-dwarf sun live longer and have higher chances for life (Superhabitable Worlds). Another new finding is that habitable worlds require sparse star environments because of frequent destructions by gamma ray bursts.
5. A "measurement" occurs when a system has to jump into an eigenstate (e.g., Stern-Gerlach mag field or polarizer). It is still inconsistent whether a detection (of new information) or a collapse is required.
6. In particle and condensed matter physics, **Goldstone bosons** (1962) or Nambu-Goldstone bosons (NGBs) are massless bosons that appear necessarily in models exhibiting spontaneous breakdown of continuous symmetries [approximate examples are phonons, magnons, pions, and goldstinos]. Exception loopholes to Goldstone theorem are superconductivity (Meissner effect) and the Higgs mechanism. Gauge fields bypass the theorem.

Summary from December, 2014:

1. **Kerr solution:** [1410.6626] took 50 years to find and even more to appreciate, only describes an asymptotic metric at Late Times after a black hole has settled down—so we still use Schwarzschild during collapse. There is not yet any interior solution. A transformation r→ r + ia cos(theta), takes Schwarzschild into Kerr. Interior Schwarszschild has Much more volume than any Euclidean estimate (e.g., Rovelli 1411.2723).

2. Zeilinger uses the name `Entangled Entanglement' for GHZ triplet states (1410.7145). GHZ (Greenberger-Horne-Zeilinger, all-up & all-down) is stronger entanglement than W states (1208.0365).
3. Half of all stars may exist <u>outside</u> of galaxies (were thrown out from galaxies) [ScienceMag 11/7/14]. This is deduced from **EBL** (infrared extragalactic background light anisotropy).
4. The Hubble constant is near Ho= 73 ± 2 for objects but 67± 2 for the CMB! [0709.3924]. A quasar has been found at z = 7 [1411.5551], and Cosmic reionization ended at z ~ 6 [1411.5375]—the last major phase change of the baryons in our universe. Einstein got Perihelion shift from First Order 1± α/r leading to an inverse cube [1411.7370]. How can that be?
5. Other: Half of the raising of ocean level from global heating is due to the <u>expansion of water.</u> Birds see in four colors: R,G,B, and near UV—and dinosaurs likely did too. And this is a big plus for appreciating the colors of feathers. Some bacteria also have an immune system and can remember viruses.

<u>Summary from November, 2014:</u>

1. Zoltan Fodor's talk on Hadron Lattice-QCD reveals much improvement so that even the n-p mass difference is now calculated (way ahead of expectation due to clever algorithms). [CU Colloquia].
2. **<u>Nobel Prize for physics 2014</u>** went for **Nakamura** (1994), Akasaki, and Amano for blue light-emitting diodes. These still need rare-earths and have efficiency droop (2007) from Auger recombination. Chemistry went to Betzig, Hell, and Moerner for finding a way to blast through microscope limits by using fluorescence to coax objects to reveal details through their own light. [predictions had included James Scott's CU work on FeRAM technology from 1989 and for organic LEDs for chemistry].
3. China is afraid to increase freedoms (e.g., present Hong Kong crisis) largely because of the <u>precedence of Mikhail Gorbachev.</u> Its economy is now at **$17.6** trillion finally exceeding the U.S.
4. Supreme Leader Ali Khamenei's dislike of the U.S. is connected to his management of Iran in the war with Iraq (US backing and chemical weapons) and because he was jailed six times and <u>tortured</u> by the Shah's secret police (who were trained by the US).
5. CU's Noah Finkelstein now teaches <u>Matter Wave</u> perspective (electron delocalized waves during propagation collapsing to particles at detection). But Copenhagen/Agnostic and Pilot-Wave is also mentioned. Students personal opinions are discussed and updated.
6. The ``sigma'' particle is intended as a quantum fluctuation of the Vacuum quark condensate or as like a ``QCD Higgs''. But present measurements focus on it being the ``fo'' scalar(near 600 MeV) which most believe to be a pion-pion resonance and maybe a tetra-quark.

<u>Summary from October, 2015:</u>

1. **<u>BOREXINO, SUN:</u>** ``For the first time in the history of scientific investigation of our star, solar energy has been measured at the very moment of its generation'' (primary seed neutrinos from p+p → d, 420 keV). Core nuclear energy to surface sun photons takes 100,000 years. This also means that the solar constant hasn't changed for 100,000 years.
2. Over the last century, Tidal Waters worldwide have risen an average of 8 inches (glacial melt and water expansion). But also the US Eastern Seaboard is sinking! (perhaps from fresh water extraction).

3. Early mammals and dinosaurs with feathers have both been pushed back to 200 million years ago. Mammals lived with dinosaurs for at least 100,000 years. Life-giving oxygen now goes back 3 billion years.
4. Some people have four functioning color cones in their eyes (blue, green and two different reds [not like birds with an extra UV color]). For an example see: Antico Concetta Artist: http://concettaantico.com/selected-works/Landscapes/
5. Europeans are made from hunter-gatherer DNA (45 kya) + early farmer DNA (~ 9 kya) + a mysterious ghost Eurasian DNA from about 4-5 kya. This ghost lineage no longer exists but is related to American Indians and a **24 kya Malta boy skeleton from Siberia**. The mystery add first appeared in Scandinavians. [SCIENCEMAG 5 Sept 2014].
6. Solid State superconductivity has examples of Higgs-like fields [ScienceMag]: a tera-Hz photon can cause a coherent oscillation of a Mexican Hat order-parameter amplitude, Φ , for niobium nitride which is detectable.
7. It is now established that artificial sweeteners can cause more weight gain than equivalent sugars. One mechanism is the alteration of gut bacteria to produce glucose intolerance/metabolic disorder. Alternately, sweeteners signals without actual energy leads to physical/mental compensation to eat more later to make up for it. There is also a possible safety issue.

Summary from September, 2014:

1. Dark Matter recent results: (1402.6703) excess gamma rays from the central milky way may indicate DM annihilations at 31-40 GeV to upsilons (b-bbar).
2. Stalin (and Czechoslovakia) provided arms to Israel in 1948 because of the hope of new socialism and the acceleration of the decline of British influence in the Middle East. Ben-Gurion headed the Marxist Mapam Party.
3. Feathers existed on dinosaurs 50 My before Archaeopteryx (Urvogel=150 Mya. Now n = 32 dinosaurs with feathers or filaments).
4. **FRB's: Fast Radio Bursts** may derive from Blitzars which may be overweight neutron stars with excessive spin which prevents them from collapsing. Over time, the spin slows, and at some point the neutron star progresses quickly into a black hole (a fraction of a second). [But new data and thinking March, 2016].
5. Digital orchestras are replacing live musicians. They use samples from real instruments, can sound equivalent, and showgoers often don't care.
6. IceCube has now observed n = 37 neutrino events from **30 TeV to 2 PeV!**
7. Biochem: In Rats: Growth Hormone FGF-1 stops diabetes. Therapeutic bacteria (E.coli Nissle) can prevent obesity. A drug (TC-2153) improves cognitive function in Alzheimers.
8. CMB was created as black-body gamma rays near z~ 2 million (and finally at z=1100 at recombination).
9. Physics Today 7/14: In 2018, a new SI will have **exact** c,h,e,k, Na, Kcd=683 lumens per watt at 555 nm. The Kg will be defined by a Watt balance. Frequency will have an exact Cesium-133 atom transition above 9 GHz (for a time standard). These in turn will make R (gas), F(charges), σ_SB exact.
10. A Kaluza Klein Cosmology (1407.7793) can simply have an extra term in its metric (1-kr^2)dψ^2 for the 5th dimension.

Summary from July, 2014:

1. **The Moon** formed after a collision of the Earth and ``**Theia''** (the Giant Impact Hypothesis is supported). We now know that the back side of the moon is heavily cratered while the front side has maria from vast lakes of basaltic lava. Why? Because the Earth and Moon used to be ~15 times closer with an early hot

Earth at 2500 C. Tidal locking made the front moon face see this heat so that a thin crust more easily punctured was formed rather than the back-side thick crust becoming cratered.

2. **EXO-200** search for neutrinoless double beta decay (and hence Majorana neutrinos) now set at >1.1E+25 years (90% confidence). But, it is located at WIPP (nuclear waste isolation pilot plant in NM) which just shut down after a fire and explosion.
3. Amazon book sales are now lower than Amazon electronics sales (and soon Amazon Grocery sales).
4. Free Will: We have an illusion of self and free will but are highly predictable in terms of demographic traits and buying habits. The brain has background **noise** which precipitates decisions independently of cause and effect.
5. Current inflationary theory is so flexible that is immune to experimental and observational tests, it is not falsifiable.
6. ‘t Hooft believes that the ``reality behind QM’’ is simple cellular automata and that there may be no need for randomness nor mysticism in quantum theory (1405.1548).
7. The Great Courses includes ``12 Essential Scientific Concepts’’ as vital bedrock ideas. One is ``String theory, membranes, and the multiverse.’’ ‼??

Summary from June, 2014:

1. Helps for old age: baby aspirin against inflammations (but questioned again in 2016), drug D-PDMP eliminates the risk of heart attack, marijuana apoptosis against cancer, antidepressants slow Alzheimers (citalopram), rejuvenation factor in blood from the young (GDF11).
2. **OAM** orbital angular momentum of light: decide if there are really `distinct but intertwined helices’ versus `twists per wavelength?’ Phase winding factor exp(imϕ). What is special is a new spatial structure to photons. Rotating electron density won’t work, electron has to be more localized to emit a photon.
3. ‘t Hooft (1405.1548) Feynman diagram particles can alternatively be described in terms of interacting fields. ``It is ok to ask what is really going on. The Einstein-Bohr debate is not over.’’
4. Understanding Russia: authoritarianism derived from `Tatar Yoke,’ lack of a Renaissance, and now Putin: nationalism, church (Moscow is the third Rome), Vladimir the Great of Kiev, opposing profanity (some declared morality), west is decadent. But mafiyas and security forces are similar (KGB veterans).
5. **Neanderthals**: NOVA. Neanderthals were in Europe for many hundreds of thousands of years. They were smart and human-like. They adapted, and their immune systems adapted. They didn’t need to be exterminated, they were just genetically swamped out during numerous interbreadings by all the modern humans about 40,000 years ago. But intermixing led to humans also acquiring beneficial genes such as improved immune systems (and more allergies). Europeans are about 3% Neanderthal, Asians about 1% and Africans 0%. Tuscans have 4%. And then there are also the Denisovans who contributed to the Australians.

Summary from April, 2014:

1. More than 10 barrels of waste are created for every barrel of oil pumped from the ground. This toxic waste is declared legal by the EPA.
2. **Heisenberg** and Dopel had the world’s first nuclear accident (6/23/42, L-IV reactor caught fire from hydrogen and exploded scattering uranium oxide.). Their

final reactor was still short of criticality (and lacked proper control rods if it had gone critical!).

3. In string theory, the original John Schwarz 1984 motivation of type-1 anomaly cancellation is now abandoned as just a red herring. The decisive success is really the personality of **Ed Witten** jumping into the arena.
4. The shortest known period star orbiting our galaxy's supermassive black hole, SO-102 with a period 11.5 years.

5. ``Broken Arrow:'' accidental A-bomb droppings: 30 kt MK-6 A-bomb hit a house in **South Carolina** in 1958 and then an H-Bomb 3.8 Mton hit North Carolina in 1961 (secondary remains buried in a swamp [U238, Li6/D, Pu spark plug]).
6. After **Cavendish's** death (1810) it was discovered from his notebooks that he anticipated: Ohm's law, capacitance, dielectric cnst, Dalton's law, Coulomb's law, mechanical thermo, and Argon gas.
7. There are Free Quantum computing sources: 700 pg Book: http://michaelnielsen.org/blog/quantum-computing-for-everyone/ , Lecture 10: http://www.scottaaronson.com/democritus/
8. A turning point in the German-Russian war was the battle of **Kursk** with info supplied by spy at Bletchley Park, John Cairncross. Without Ultra, Hitler would have won and carved up Russia.
9. L and S do not commute with H, so there is no agreed upon relativistic spin operator.
10. Feynman believed that not everything can be or should be reduced to an explanation of just a few minutes and a simple but potentially misleading real-world analogy. (a problem that we constantly deal with in our more popular science readings)

AND MANY MORE NOTES GOING BACK DECADES….. dp