A Stroll Through Physics

David L. Peterson

Dedicated to My Family: Faye, Shayna, Lisa

And to Mike Jones [*Dr. R. Michael Jones, Boulder*]
For four decades of great physics discussions.

Dave Peterson, 4350 Drew Circle, Boulder, Colorado, 80305
Email: davepeterson137@gmail.com
Date: May 9, 2014

CONTENTS:

May 9, 2014, dp

"A Stroll Through Physics.pdf''

<u>NEW REPORTS > 2010</u> <u>[LaTeX]:</u>

``General Relativity for Pedestrians,'' First order approximation, 1974, DP.
THE GENERAL THEORY OF RELATIVITY, 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

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

Sample First Pages of Key STK Publications [D.L. Peterson]:

`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

Not Shown: *STK Symposium*:
`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.

A STROLL THROUGH PHYSICS

DAVID L. PETERSON

Abstract. The following is a brief summary and a few highlights of my discoveries and thoughts on the workings of Nature along with some comments about a detour that we are all obliged to make to exist in the practical world. Although difficult, it is my view that physics offers the best way to know the basic fundamentals of Nature. It is uplifting and provides natural substance and purpose for everyday existence. Although some of the work here is original, my primary interest is in presenting modern physics (particles and fields) at an intermediate level between mathematical rigor and heuristic talk. We wish to understand physics clearly, but fundamental material is difficult to find in texts or even in journals.

"We shall never cease to stand like curious children before the great mystery into which we were born." [Albert Einstein]

"I was like a boy playing on the sea-shore, and diverting myself now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me." [Isaac Newton]

"The effort to understand the universe is one of the very few things that lifts human life above the level of farce." [Steven Weinberg]

1. Background

I love physics! It covers an incredible range from smallest to biggest − from elementary particles to hadrons to nuclei to atoms and all the way up to the universe as a whole (and some believe even beyond that). It tries to explain everything from the most basic fundamental principles of Nature. It uncovers ever more reliable truths with its dovetailing of theory and experiments. It reveals an amazing mind-boggling reality that pushes our imaginations to greater and greater heights with each new generation. I love its history! I want to experience and re-live those greatest "Ah-Hah!" moments when new theories and discoveries fell into place. I love sharing and teaching those beautiful ideas I treasure most. I crave keeping up with the latest developments in physics and much of science in general.

Date: December 23, 2011.

email: davepeterson137@gmail.com. Paper updated to 7 May 2014.

I can't get enough of the unexpected ways that the broad world makes sense. I've done this pretty much my whole life. The following is a brief sketch of what I've learned – my "stroll through physics."

I was raised in a small town in south-east Kansas named "Parsons" and noted for its railroads and wheat. My interest in science came from finding an unused Gilbert chemistry set in a neighbor's back-alley trash-can and doing all of the experiments in its guide-book. Chemical reactions were so fascinating that I would then sneak upstairs from the children's library into the adult portion of our town's big Carnegie Library to go through chemistry books. My after-school job enabled me to set up a well stocked chemistry lab in my folks garage where I could experiment to my heart's content, make my own rockets, and perform electrical experiments. Parsons was also blessed with its reptile life as a merging of north-east-south and west varieties. Everyday nature walks would usually yield a new short-term pet snake, lizard or turtle; and I ended up having over a hundred different kind of snakes for pets. Going through the local scouting program also encouraged nature studies; and I eventually became an Eagle Scout. I also built a 6-inch reflecting telescope and often looked at the stars and planets at night. As Albert Einstein once said,"There was this huge world out there, independent of us human beings and standing before us like a great, eternal riddle, at least partly accessible to our inspection and thought. The contemplation of that world beckoned like a liberation."

When I was twelve while vacationing in Montana, I found a life-changing book for sale called, The World as I see It, by Einstein. I had no idea at that time that adult humans could be wise and soon became a convert to physics. I then went through the popular science encyclopedia and a sourcebook on atomic energy. A high-school course on chemistry somehow dampened my interest in that field which wasn't rekindled until I could again do free-style experimentation for manufacturing problems when I became an engineer later in life. I was also head of a physics club in high school and got to talk about relativity. My folks moved to Colorado when I was fifteen, and I enrolled at the University of Colorado in Boulder in 1960 when I was seventeen with a major in Engineering Physics and an initial goal of doing nuclear physics. The closest I got to that goal was two summer NSF grants at the CU Cyclotron doing elastic scattering of 28 MeV protons from a deuterium gas target. The result was wiggly plots of differential cross sections showing that quantum mechanics was at work rather than classical or electrostatic scatterings. After that, my interest was drifting towards "elementary" particle physics as more glamorous, fundamental and mysterious. My favorite undergraduate course was Leighton's Principles of Modern Physics for seniors and taught by Rodman Smythe. He was tall and thin and wore baggy white shirts, and I started wearing baggy white shirts too.

Graduate training continued again at the University of Colorado with all of the standard courses: electrodynamics, mechanics, quantum mechanics, statistical mechanics, spectra, mathematical physics, advanced quantum, nuclear physics, ... but also general relativity and biophysics. I finally got to sample particle physics with a summer program on neutrino scattering at the Argonne National Labs 12.5 GeV accelerator (ZGS) near Chicago and also studied "pions on propane" at CU Boulder under Leona Marshall Libby using some old scattering photo-film-tapes from Jack Steinburger . I passed written comps and orals and got my MS in physics in 1968, but stopped there because something was amiss... too many particles, too many particle tracks, dubious theories, and a pulling-back in government funding for particle physics [1]. So, I started new graduate programs in Biophysics at the CU medical center in Denver and then the department of Mathematics at CU Boulder. The courses were interesting: biochemistry, organic, macromolecules, bacteria-genetics... and then real analysis, probability, linear algebra, complex variables, topology.... Preparing for comprehensive exams in three separate graduate departments is a bit of a stretch. And, in addition, by then I was broke and also couldn't see any bright lights at the end of the tunnels. I didn't really want to specialize, I wanted to explore all of science. The Viet-Nam War was in full swing, and jobs were scarce. It doesn't take much to be a perpetual student, a janitor's pay can do it. But ultimately, one has to settle down and be practical; and with my background I could finally do that.

2. Research Interests

I've always had an interest in teaching and a desire to simplify key physical concepts for a wider audience and my own clearer understanding. Fortunately, I was able to teach physics as a teaching assistant in the University of Colorado Physics Department for several years from 1965 and then again in the department of Mathematics from 1970. I also taught and tutored and wrote courses for continuing education and minority programs during the seventies until 1981. Rather than just teaching standard content, I wanted to stress the mystery of the underlying physics and motivation for the student to dig further. I always wanted to address, "What is really going on underneath the words and the mathematics? What is the Physics?" There is a goal that might not always be satisfied that if something is really understood, its essence should be explainable to a patient high school student.

If a goal is to explain things simply, then why are the following reports so mathematical? As Feynman said, math is not just another language, it is language with logical reasoning built into it. Physics is a science of relationships, and math makes those relationships not just crisp and clear but also quantifiable for comparison with numerical measurement. A set of observations might be reduced to a simple formula; and while formulas should be expressible in words or possibly in pictures, a useful formula is worth a thousand words. For presenting clearly, we want just the right level of math, a Goldilocks level. Excessive mathematical detail may confuse. And, as we used to joke in math, excessive rigor can

[1]1968 marked the beginning of high inflation rates followed by the discouraging of non-military pure research funding. And, new clarity in high energy physics really had to wait until about 1974 with the "November Revolution:" the discovery of the charm quark, and the arrival of the "Standard Model" of particle physics

lead to rigor mortis.

One early interest was in the foundations of physics and somehow simplifying general relativity (GRT) for general understanding for non-specialists. I had a fascinating graduate class in 1967 under Ron Adler[1] which used the traditional tensor notation. But I later saw an article from Leonard Schiff which made some of Einstein's tests as simple as special relativity combined with the principle of equivalence. Ordinary Newtonian gravitation is merely the first order curvature of time (using $g_{oo} \sim 1 + h_{oo}$). Bending of starlight is half due to space curvature and half to time curvature (metric terms g_{oo} and g_{rr}). I used this as a basis to simplify much of standard general relativity and wrote it up along with my own derivations in papers beginning in 1974 and evolving into a long book of typed studies in the general theory of relativity (or "GRT") [3].

I then made up my own thesis, "Can the perihelion shift of the planet Mercury also be decomposed into intuitively simple principles?" It was well known, for example, that special relativity by itself gives one-sixth of the Einstein shift per century. Since a gravitational field itself might be considered to add negative energy (and hence mass) feedback to an inverse square field, could it also contribute to a perturbation of orbit? It turned out that "mass-feedback" was not an invariant concept. But, worse, it also turned out that the decomposition of perihelion shift in terms of bending of time and space depended on the choice of metric (it also wasn't invariant − the "Decomposition" paper by me listed below). The common approach is to use a PPN Isotropic metric [2] which has a first order space term $2\gamma m/r$ but two time terms $-2\alpha m/r + 2\beta(m/r)^2$ giving an elliptical shift as proportional to $(2\alpha - \beta + 2\gamma)$. But a lesser known Schwarzschild form without any second order β term gives a shift proportional to $2\alpha + \gamma$. If I had selected this topic for a thesis, it wouldn't have held up. On the other hand, the perihelion shift can be explained without using the full power and complexity of general relativity.

But this raised another interest, "Can the Isotropic metric versus its Schwarzschild form be conveniently pictured for the standard cosmologies, S^3, E^3, H^3 ?" Nearly all texts rely on abstract mathematics without visualization. This topic was answered in my other home research work ("Graphical Representations" listed below). I worked this out in 1979 but also left it unpublished.[3] Interest in General Relativity was kept active by periodic meetings and discussions with Mike Jones − and we continued meeting from the mid 70's until today (2014). He wrote up a large book on his studies, "Quantum Gravity and Mach's Principle;" and I wrote up a 240 page proto-manuscript on general relativity (never published). At about this time (the 1970's), the momentum of research and interest in GRT transformed from something of a backwater into fast moving and larger scale mainstream

[2]for source mass $m = MG/c^2$, isotropic radius r, and PPN parameters α, β, γ.

[3]The first two papers following this introduction are these topics in general relativity. They are somewhat difficult and advanced, and my early thinking was that they were original. I was fairly proud of them, but experts in the field might not agree. The articles following them are easier but still usually require some math background.

physics − leaving us substantially behind.

Cosmology too has been difficult and abstract. I simplified much of it in a way suitable and understandable for early physics studies [above papers on The Radius of the Universe and Cosmological Distances]. A great many of my discoveries are now just common fare for early courses in cosmology − for example in Mark Whittle's Great Courses presentations [4]. Going through that course was an amazing eye-opening experience with its intricate dovetailing and consistent story telling of the now "Standard Model" of cosmology. But self-struggle and my personal discoveries aided my understanding and appreciation for the subject. You get out of something what you put into it.

I worked my way through college and had more than fifty jobs. But my first long term job was as in engineering. From 1978 to 2009, I worked at Storage Technology Corporation in Louisville, Colorado as an engineer and covered many of the engineering titles (manufacturing engineer, test, quality, standards, systems modeling, development, mathematical modeling, and systems engineering). I always kept up with physics journals as a side activity − mainly the realm of particles and fields. The primary focus of my engineering arena was on magnetic recording for high-capacity hard-disk and tape drive data storage. Since I had a degree in Engineering Physics and since physics stresses problem solving and mathematical modeling, I just continued to use these over a broad range of general physics (electricity and magnetism, mechanics, fluids, diffusion, thin film optics, spectroscopy, chemistry, acoustics, and performance modeling). Manufacturing had a lot of great physics problems (unfortunately, it has now largely been outsourced). Apart from standard bureaucracy, I got to apply my knowledge, grow, had a good time, had five issued patents, and wrote thousands of reports (most of which would be considered company proprietary). I was also fortunate in solving two of the biggest problems in the industry: loss of surface spinning-disk lubrication from years of wind shear stress and loss of recorded data due to micro-contamination from rare-earth magnetic particles (and my papers on these were published in journals). Many of my other later papers were on the discovery of fractal data patterns or power laws in disk, tape, and cache storage and acquisition. I mainly contributed to the IEEE Transactions on Magnetics and the Computer Measuring Group Proceedings (CMG).

Technical developments are often very short-lived: proud today, gone tomorrow. I saw a picture in 2008 comparing our heavy 100 pound disk drive (just the head-disk assembly portion) against a tiny flash memory the size of a postage stamp. Both held a giga-byte of data. Later hard disk drives got up to the tera-byte (TB) range in capacity and beyond, but some new PC's have only high capacity flash memory without any hard drive. We were so proud of our technology, but it is now quite a thing of the past. And our nine-building site in Louisville, Colorado was totally razed for a proposed Conoco-Phillips research facility (which was then blocked by the deep recession of 2009).

One perpetually interesting question was, "What really is a magnetic field?" Since we were in the magnetic recording business and the field is nearly a century old, it would seem that someone should know − but I never got a satisfactory answer. My only conclusion was that $F = qv \times B$ is essentially a "Coriolis force" due to being in the wrong frame of reference when electrical currents are present. But this view is probably not found in any text-book. It is well known that effective momentum in classical electromagnetism, special relativity, and quantum mechanics is given by $p' \simeq mv - eA$ where A is the vector potential (and this potential is being increasingly appreciated − especially since the Aharonov-Bohm effect for electron interference fringe shifting). A question now is, "in what sense can $\vec{A}$ be considered like a fluid flow?" and does internal symmetry space involve higher dimensions? (like the 5-th dimension of Kaluza-Klein or a section of string theory).

Much of physics can be understood in fairly elementary ways. A giant exception is quantum mechanics (QM) which represents a very different world from classical physics and is generally presented only in abstract form. There are quite a few mutually incompatible "interpretations" for quantum mechanics with no general agreement among physicists. The early Copenhagen Interpretation was dominant from 1935 to nearly the modern times − almost to the point of being dogma. Now that key experiments have been performed clearly demonstrating "entanglement," people are appropriately confused. I had the (misfortune?) of reading David Bohm's book, Causality and Chance in Modern Physics, back in 1962 before my first class in quantum mechanics. I've been puzzled and disappointed ever since by the overall lack of interest in trying to make sense of it and the dogmatic view that one really shouldn't try. Now that I am retired, I can continue my struggle to understand QM. That is the topic of my lengthy unfinished report, "Beneath Quantum Mechanics" followed by a later study of "Sub-Quantum Physics." It is understood that I am highly unlikely to make a real breakthrough in this topic; but one positive aspect to that is that this project will never end either. There is an old Russian toast, "To the success of our hopeless task." [4]

Looking daily at online "ArXiv.org" (meaning archives) under quantum mechanics makes it clear that there are now a great many interpretations with diminishing consensus. It is now acceptable to change our minds about which if any interpretation is most probable. Since the 1980's, I've been fond of parts of John Cramer's "Transactional Interpretation" − at least for the idea of back-and-forth in time communications between sources and detectors. This view is still just a minority, but I can't see how to avoid it. It sure helps understanding of mechanisms for contextualities and entanglements. Aharonov's "Two-State-Vector-Formulation" has some similarities to Cramer, and interest in TSVF and in weak measurements seems to be growing. The fundamental laws of physics are time invariant, so thinking of time going backward in the microworld shouldn't be so strange. The

[4] "Za uspyekha nashevo beznadyozhnovo diela." This was made famous in biographies on dissident physicist Andrei Sakharov while exiled away in Gorky with his companion Elena Bonner. It was also used by the upper Soviet military while in Afghanistan. Sakharov actually achieved success for awhile under Gorbachev, but no one wins in Afghanistan.

big mystery is why in the classical world time only goes forward.
A primary conceptual difficulty in quantum mechanics has been the apparent existence of particle properties (such as charge or magnetic moment) for single particles in multiple paths at the same time. But it now seems clear that the quantum fields existing everywhere in space-time already contain this information so that it is accessible everywhere anytime. It doesn't have to be carried by a wave function. I don't yet have the answers to the hidden mechanisms of quantum mechanics − I don't think anyone does.

I also strongly wish to understand the substance of the "vacuum of space-time." A popular picture (based on string theory) is the existence of six-tiny curled up dimensions at each point of classical space-time. The Kaluza-Klein and Gauge Theory models suggest that physics may depend on phases existing on these curled up objects (e.g., Calabi-Yau spaces) so that E&M is like a U(1) circle and perhaps QM phase is also another little circle with phase going around it. I am writing and thinking deeply about these things, but they are pretty intangible. The trend of accepting the existence of physics at the Planck scale still seems unlikely to me. And I wish a "theory of everything" to explain quantum mechanics rather than simply assuming it.

There is an evening club in Boulder consisting mainly of old-timers (like me) and focusing on finding intuitive understandings of cosmology and particle physics. Book discussions have encouraged me to write up my understandings (partial list included here) on basic topics such as inflation, general cosmology, general relativity, astrophysics, the standard model, gauge theory, basic quantum mechanics, black holes and Hawking radiation.

3. New Reports, D. L. Peterson

One of my goals since 2010 has been a summarization of views and understandings in LaTeX.pdf document form. These fall roughly into two categories: Cosmology/Relativity and Quantum/Particle physics:

- **"Graphical Representation of Radial Coordinates in Cosmology, "** DP, 10 January, 2011, 6 pages. Graphical Representation.txt.
- **"Decomposition of the Perihelion Shift of General Relativity,** DP, 31 January, 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, to 5/14/11 (10/5/11, Cosmology_1.tex).
- "Comments on Inflation," DP, 11/10/2011, 9 pgs. .pdf
- "Hawking Radiation," 12 pgs, 9/6/13.

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

- "The Last Decade in Experimental Particle Physics" updated to 2/9/12 11 pgs.
- "Rotations of Base States.pdf," 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.pdf
- "Explaining S Orbitals and bonding," DP, 6/24/2012 11 pgs. .pdf to 3/6/13.
- "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, 30 pages.
- "Gauge Theory for Electromagnetism," DP, 26 pages, November 6, 2011, 12/22/11 . pdf.
- "Five Dimensional View of Electricity and Magnetism," 9/15/05, 2010, 8 pgs.
- "Circle Models in Modern Physics," 3/6/2012, 10 pgs. .pdf.
- " Special Topics", water waves, matter wave index 8/21/12.

4. Comments on Papers:

Graphical Representations: Many of us wish to have pictorial representations of the mathematics that we use. I still haven't seen any texts on general relativity that attempt to show visual representations of the important isotropic radial coordinate either for the spherical or hyperbolic Robertson-Walker universe. My picture interpretations would have been a welcome addition. Of course, now, we believe that we live in a flat Euclidean universe.

Decomposition: Most discussions of the general relativistic "parameterized post-Newtonian (PPN) line elements use the isotropic metric form seemingly preferred by astronomers. The Robertson and Noonan text is a rare exception and uses a Schwarzschild form line element. These are related by diffeomorphism but result in different perceived contribution of the perihelion shift for space and time. Although Newtonian gravitation sees the gravity field as having negative energy density as a function of radius from an attracting body, this is not a legitimate contribution to the GR perihelion shift of Mercury. The Principle of General Covariance also does not apply here.

Radius of Universe: It is interesting that elementary Newtonian arguments can be used to approximate the results of general relativistic cosmology. The Einstein-DeSitter model (EdS) was a preferred cosmology from 1932 until the discovery of dark energy (1990s). EdS is a matter-dominated Friedmann model with zero curvature which has just the right amount of energy for a universe to escape to infinity. The Einstein equations are needed when curvature $k \neq 0$.

Cosmological Distances: A variety of different types of distance measures are used in cosmology. They are discussed and defined poorly in most of the available literature and can easily confuse the student. This paper demonstrates these measures and also discusses calculation of these distances for a ΛCDM universe.

Inflation: The idea that the early universe underwent a period of inflation dates from about 1980 (Guth). Studies of the cosmic background microwave radiation (CMB) solidified belief in simple faster-than-light inflation models prior to the start of the Big Bang universe. Now, the recent discovery of B-mode polarization at the South Pole points towards Andrei Linde's elementary parabolic model of the inflaton.

Hawking Radiation. I was reluctant to research this area because I've always been suspicious of the idea of (intrinsically unmeasurable) radiation from the horizon. After presenting straightforward math in this paper, I am still suspicious. Worse than that, just after our cosmology book club studied it, the topic imploded on itself with the AMPS "firewall paradox" along with new stringy Fuzzballs. Hawking had intended his black hole entropy formula be used as his obituary, but now the experts themselves are in strong disagreement. A quantum theory of gravity might be required for resolution.

Recent Astrophysics: This is an update on precise parameters for the CMB, new inflation, AMS positrons at the space-station searching for dark matter, new galaxy results, the possibility of warm dark matter universe (WDM), Milky Way giant Fermi bubbles (probably from our black hole Sgr A^*).

The Last Decade in Experimental Particle Physics: Most texts and popular books are not up to date, so I gathered journal articles to present more current status (the discovery of the Higgs at 125 GeV, neutrino oscillations, quark gluon plasma, CP violation, and what is to come next. This paper was vetted last year by the University of California, Irvine.

Base State Rotations. Weak interactions effectively see the world through "twisted glasses (the Cabibbo angle, the Weinberg angle, the CKM V matrix for quark superpositions, and the PMNS mixing matrix for neutrinos). Instead of referring to e, mu and tau electroweak neutrino eigenstates, the mass eigenstates are different and are called 1,2 and 3.

Fine Structure Constant, alpha [5], refers to the coupling constant attached to each vertex of a Feynman diagram for quantum electrodynamics (QED), $g_e = \sqrt{4\pi\alpha}$. This is simply proportional to electron charge e, yet various scattering formulas typically scale with e^4 (two couplings for two vertices per diagram plus the Born rule squaring for macro-probabilities). Numerical values are confusing to first time students because there are so

[5] $\alpha = e^2/4\pi\epsilon_o\hbar c = 1/137$ (in SI units)

many different conventions for choices of units.

Electron Spin/SU(2): Electron spin has a 2-dimensional Hilbert space, just $|up\rangle$ and $|down\rangle$. Yet we can measure spins in many directions: up and down, left and right, and in and out (of the paper) for directions z, x, and y (along with arbitrary directions). When we express y-spin in terms of z-spin, we are forced to use complex numbers. The mathematics of spin and SU(2) is strange and fascinating. The generators of the Lie group SU(2) are the quaternions (hyper-complex numbers which can also be written in terms of the Pauli matrices, $q_i = \pm i\sigma_i$.)

Bonding with S-orbitals: I believe that introductory chemistry is almost always taught badly. It often starts with Lewis electron sharing dot notation (pre-QM) that may satisfy some practical students but is horribly unphysical. Almost anyone who desires to dig deeper is guaranteed to become confused. Part of the problem is fear by writers to express underlying reality because it depends on interpretations of quantum mechanics: are we dealing with electron particles at all or just waves (etc.)? One solution is to bring the Born rule to the front and discuss bonding as an enhancement of orbital overlaps. This paper discusses puzzles and possibilities for underlying clarity.

Aspect Experiment: Alain Aspect's 1982 experiment was the first careful test of violation of the Bell inequalities showing that quantum reality is non-local (entangled with apparent faster than light communications). Prior to this, quantum mechanics was dogmatically Copenhagen (despite jabs by Schrödinger, Einstein, Bohm, and Bell). After this, the floodgates of multiple interpretations began.

Beneath QM: This 43 page paper ventures guesses at the reality beneath quantum mechanics as an holistic quantum communication network for each quantum field excitation. In terms of labels, there must now be over a hundred different interpretations of the mathematics of non-relativistic quantum mechanics with no consensus in sight. I appreciate portions of the "Transactional Interpretation" and believe that communication occurs both backwards and forwards in time along the network (e.g., entangled world lines). I am also fond of Wilczek's GRID as a set of quantum fields Platonically occupying all of space-time. I believe that the Born rule implies that QM should be viewed as a square root of classical reality thus requiring complex numbers (and quaternions and Clifford algebras).

Sub-Quanta Information: This 32 page paper says that a psi-wave is essentially a physical code enabling knowledge of momentum and energy. Electric charge need not be carried by psi because knowledge of e is contained in the electron quantum field occupying all of space-time. QM is about waves and fields rather than so called particles. The wave-function of single photons is expressed by Maxwell vector equations. The Vacuum is a mathematical machine. Composite particles are wavicles about wavicles about any selected center of energy.

Gauge Theory: A common claim from books is that local gauge invariance yields Maxwells equations, but they rarely fully justify this claim. It has to be done with relativistic Dirac Lagrangians rather than the non-relativistic Schrodinger equation. Discussion needs to include something resembling "real" (non-integrable) vector potentials $\vec{A}$ above and beyond the unreal "del chi" ($\nabla\chi$ gradient of scalar fields) which just go along for the ride. One of the few texts to do this right is Aitchison and Hey. This is a very tricky subject requiring careful use of the various gauge definitions.

Five Dimensions: Most discussions of 5-D Kaluza-Klein theory (KK) are horribly opaque and only show the math. Does KK have any real heuristic use for electromagnetism? The standard claim is that it yields Maxwells equations for E&M, but this is only true of the source-free Maxwell equations. KK is a stepping stone to string theories. Here I use it to discuss "Circle functions in modern physics" primarily for electromagnetism.

Nuclear Shells: One of my first submitted papers was a special consistent pattern I discovered for nuclear shell filling. I still believe that this and the standard electron shell filling scheme for chemistry should appear in many texts − but my idea hasn't shown up anywhere. I was so proud of it that I once presented my paper to Linus Pauling when he visited CU − like a kid giving a new pet turtle to his mother.

An Interesting Function: I was once a grad student in mathematics and enjoyed teaching it to students. One simple heuristic paper was on the interaction portion of the energy density of E fields from two point charges, $\rho = \epsilon_0 \vec{E}_1 \cdot \vec{E}_2$. Apart from it's fascinating "African Mask" contour profile, it has singularities at the charges which have value 0 from the side, plus infinity when approaching from the back, and minus infinity from the front (three different limits depending on the view).

Pre-2010: Other discussions include a long project on presenting general relativity in elementary and intermediate ways.

Some other projects for engineering work were almost PhD level in their novelty and duration. The one I liked the most was on the gradual depletion of lubrication from the surface of rotating magnetic hard disk drives. It was universally believed that this was due to centrifugal force from fast spin; but I showed that it was instead due to wind shear stress. Other novel features were gradual proportional replenishment from subsurface binder and failure of the commonly believed "no-slip" condition of surface lubrication. Ten years of testing included use of material's lab ESCA surface lube thickness versus radius, FTIR for total lube thickness, and "annular Freon strip-and-weighs" showing migration from annuli to disk rim. There was also a direction of migration depending on surface thickness.

Power Laws: Another project was a search for previously unknown statistical patterns in data cache. I had six publications showing that fractal patterns or self-similarity existed in DASD(direct access storage devices) I/O traffic, DASD cache fast memory buffers, track skew and track inter-reference times, and duration of open files. Power law (fractal) distributions also pertained to robotic tape: sizes of files transferred, catalog files, dataset accesses and interarrival times. These aid in understanding and modeling of cache hit and miss rates. There are many citations to these "heavy-tail" publications in computer journals. Power laws often emerge from complex systems.

I was also a member of the magnetics society (IEEE Transactions on Magnetics) and did a lot of magnetics modeling for magnetic data storage devices. I essentially had a piece-wise model suitable for calculations prior to the use of MR recording heads. Magnetic modeling also came in handy for calculating laser scribed disk error standards and effects of microscopic super-magnetic contamination on data. Magnetics and magnets are fun.

References

[1] Ronald Adler, Maurice Bazin, Menahem Schiffer, Introduction to General Relativity, McGraw-Hill, 1965. Used for Physics 695 in 1967.

[2] L.I. Schiff, "On Experimental Tests of the General Theory of Relativity," American Journal of Physics Vol. 28, No. 4, 1960 p 340.

[3] David L. Peterson, "General Relativity for Pedestrians," 1974, unpublished. Portions also found in my "General Physics 201-4" for CU Lifelong Learning, Independent Study Syllabus, 1977. These evolved into a book of writings, "An Elementary Introduction to the General Theory of Relativity," 1977-2011, 220 pages (unpublished manuscript).

[4] Mark Whittle, Cosmology: The History and Nature of our Universe, The Great Courses Guidebook, The Teaching Company, 2008. This is a wonderful DVD series that I highly recommend to everyone.

GRAPHICAL REPRESENTATION OF RADIAL COORDINATES IN COSMOLOGY

DAVID L. PETERSON

ABSTRACT. The isotropic radius for the spherical S^3 Robertson-Walker cosmology at a hyperpolar angle χ_o is the length of the line segment of the tangent to the circle at the 'bisector' $\chi_o/2$ bounded by rays from the 'origin' to $\chi = 0$ and $\chi = \chi_o$. A similar concept applies to the case of the isotropic radius for the hyperbolic H^3 universe except that Minkowskian distances are now required.

The exact exterior solution of Einstein's equations for general relativity for a non-rotating spherical central mass is given by the Schwarzschild metric

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -(1 - \frac{2m}{r})c^2dt^2 + \frac{dr^2}{(1 - \frac{2m}{r})} + r^2d\theta^2 + r^2\sin^2\theta d\phi^2 \tag{1}$$

where $m = MG/c^2$ (and a frequent convention is to set $c \equiv 1$ and $G \equiv 1$). This uses Schwarzschild radial coordinate r and has a Schwarzschild coordinate singularity at $r = 2m$. One way to visualize Schwarzschild geometry in Schwarzschild coordinates is by constructing the"Flamm Paraboloid" [1] with an added embedding parameter $z = z(r) = \sqrt{8m(r - 2m)} + cnst$. The metric (1) can be transformed into a more Euclidean form through change of variable for the radial coordinate, $r = \rho(1 + m/2\rho)^2$.

$$ds^2 = -\frac{(1 - \frac{m}{2\rho})^2}{(1 + \frac{m}{2\rho})^2}c^2dt^2 + (1 + \frac{m}{2\rho})^4d\ell^2 \tag{2}$$

The space term $d\ell^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2d\Omega^2$ where $d\Omega^2 = d\theta^2 + sin^2\theta d\phi^2$ is an element of solid angle. This metric $eqn.$(2) is called the "isotropic" metric or the "isotropic Schwarzschild" line element to distinguish it from other isotropic metrics such as that used for cosmology. An isotropic metric form has the virtue that the relativistic local curved space coordinates look more like Euclidean classical coordinates.

Two previously traditional candidates for the large-scale spatial structure of the universe were the hypersphere, S^3, and hyperbolic space, H^3. This applies to an isotropic and homogeneous universe with constant curvature and without cosmological constant. In addition to the now preferred zero-curvature Euclidean space, E^3, these geometries are Robertson-Walker solutions of the Einstein Field Equations of general relativity representing a closed universe of constant positive curvature and an open universe of constant

Date: 10 January, 2011.

negative curvature. There are many different metric forms for the spatial line-elements for these geometries, but the most common ones are the angular (or polar Gaussian) metric, the Schwarzschild form (or curvature) metric, and the isotropic metric. [1]

$$S^3: \quad d\sigma^2 = a^2[d\chi^2 + sin^2\chi d\Omega^2] = \frac{dr^2}{1 - r^2/a^2} + r^2 d\Omega^2 = \frac{d\rho^2 + \rho^2 d\Omega^2}{(1 + \rho^2/4a^2)^2} \tag{3}$$

$$H^3: \quad d\sigma^2 = a^2[d\chi^2 + sinh^2\chi d\Omega^2] = \frac{dr^2}{1 + r^2/a^2} + r^2 d\Omega^2 = \frac{d\rho^2 + \rho^2 d\Omega^2}{(1 - \rho^2/4a^2)^2} \tag{4}$$

The varible "a'' = a(t) is called the "cosmic scale factor" (or for S^3, the "radius of the model universe"), "χ" is the "hyperpolar angle," r is the Schwarzschild type radial coordinate, and "ρ" is the "isotropic radius." The radial coordinates essentially include the expansion factor over time (e.g., $x = a(t)\sin\chi\cos\theta\cos\phi$). In the original static Einstein spherical S^3 universe, the g_{rr} term was $(1 - r^2/R^2)^{-1} \simeq (1 - \Lambda r^2)^{-1}$ where Λ is the "Cosmological Constant." [2] The Λ concept was then dismissed as a "blunder" but is now re-appearing due to the dominance of universal "dark energy."

The angular forms for the hypersphere and hyperboloid hypersurfaces have been graphically shown [1]. And it is known that the Schwarzschild radius coordinate is the perpendicular distance from an embedding axis v for $\chi = 0$. That is, in Figure 1 below, the cosine horizontal projection of small arc length gives: $ad\chi = dr/\cos\chi = dr/\sqrt{1 - \sin^2\chi} = dr/\sqrt{1 - (r/a)^2}$, as in the metric expression. Similarly, for H^3 where $r = a\sinh\chi$ in Figure 2, note that Minkowski arc-length is given by:

$$(ad\chi)^2 = \frac{a^2\cosh^2\chi d\chi^2}{(1 + \sinh^2\chi)} = \frac{dr^2}{(1 + (\frac{r}{a})^2)}$$

and the horizontal projection of length $ad\chi$ is given by $\cosh\chi$, i.e., $dr = a\cosh\chi\ d\chi = (\cosh\chi)(ad\chi)$. However, the representations for the isotropic radius are not well known and will be shown here.

Once one metric form is known, transforming from that form to other forms is a straightforward exercise. For example, matching the coefficients in the expression

$$d\sigma^2 = a^2[d\chi^2 + sin^2\chi d\Omega^2] = \frac{d\rho^2 + \rho^2 d\Omega^2}{A(\rho)} \tag{5}$$

gives $a^2 sin^2\chi = \rho^2/A(\rho)$ and $a^2 d\chi^2 = d\rho^2/A(\rho)$ so that $a^2 A(\rho) = \rho^2/sin^2\chi = d\rho^2/d\chi^2$. Integrating the expression $d\chi/sin\chi = d\rho/\rho$ and using the initial condition A(0) = 1 and considering small angles yields an expression for 'change of variable' (χ to ρ) and result for A(ρ):

$$\tan(\chi/2) = \rho/2a, \quad and \quad A(\rho) = [1 + (\rho/2a)^2]^2 \tag{6}$$

for S^3. A useful intermediate step is: $\sin 2\frac{\theta}{2} = 2\tan\frac{\theta}{2}/[1+\tan^2\frac{\theta}{2}]$. If the factor A is written as $A(\rho) = [1+k(\rho/2a)^2]^2$, then the factor $k = +1$ for the positive curvature of the spherical universe. As $\chi \to \pi, \rho \to \infty$.

Similar equations for H^3 give:

$$\tanh(\chi/2) = \rho/2a, \quad \textit{and} \quad A(\rho) = [1 - (\rho/2a)^2]^2 \tag{7}$$

In this case, $\int \frac{d\chi}{\sinh\chi} = \ln \mid \tanh\frac{\chi}{2} \mid$, and a useful intermediate step is $\sinh(2\frac{\chi}{2}) = 2\sinh\frac{\chi}{2}\cosh\frac{\chi}{2} = \frac{2\tanh\frac{\chi}{2}}{[1-\tanh^2\frac{\chi}{2}]}$. Again, if $A(\rho)$ is written as $[1 + k(\rho/2a)^2]^2$, then the factor $k = -1$ represents the negative curvature of the hyperbolic universe.
The key to a graphic representation of the isotropic radial coordinate ρ lies in the equations for ρ versus χ, (6) and (7).

Using a higher-dimensional Euclidean embedding variable, v, the 3-sphere can be defined as the set of points satisfying $x^2+y^2+z^2+v^2 = a^2$. For simplicity, let y = z = 0, and consider a circular arc in the x, v plane. Draw a tangent line to the arc at some hyperpolar angle $\chi_0/2$ and bound the line by rays at angles $\chi = \chi_0$ and $\chi = 0$. The length of the isotropic radius at angle χ_0 is just the length of the tangent segment: $\rho = 2a\tan(\chi_0/2) = 2\ell$. Figure 1 shows how to picture the isotropic radius. The angle of tilt, α, of the tangent line is just $\chi_0/2$. For small hyperpolar angles, $\rho = 2a\tan(\chi_0/2) \simeq 2a\chi_0/2 = a\chi_0$. A similar isotropic metric applies to the more familiar case of a 2-sphere S^2 (e.g., a basketball of radius a) using just the polar angle θ instead of the hyperpolar angle χ. The same factor $A(\rho)$ applies. The Schwarzschild radius, r, is shown as the perpendicular distance to the polar axis.

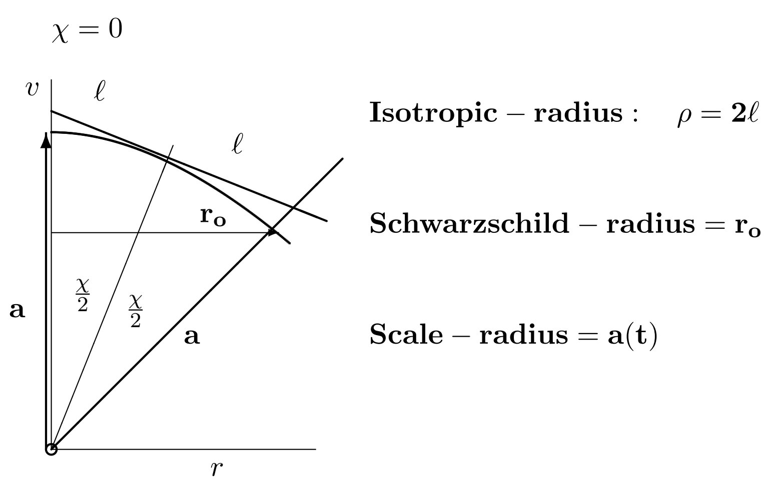


Figure 1: Radius comparison for 3-Sphere Universe at hyperpolar angle $\chi = \chi_o$.

There is a rich history of hyperbolic or Lobachevskian geometries, e.g., [5] or [6]. But A graphic representation of the isotropic radius for hyperbolic space is difficult because the negative curvature hyperboloid cannot be embedded in Euclidean space but can be embedded in Minkowski space. In particular, David Hilbert showed that there is no four times differentiable embedding of the hyperbolic plane in E^3. [4] If an embedding represented hyperbolic lengths faithfull by Euclidean lengths, the embedding is called "isometric." It is impossible to embed the entire hyperbolic plane isometrically in Euclidean space, but isometric embeddings of portions are possible in E^3.

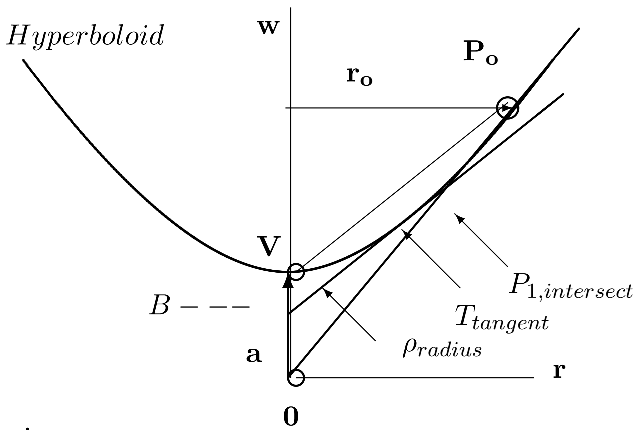


Figure 2: Radius comparison for 3-Hyperbola Universe in Minkowski Space. The isotropic radius ρ is the Minkowski length of the segment BP_1 passing through the tangent point, T.

Figure 2 shows how to identify the isometric radius for H^3. [7]. Let the Minkowski embedding axis be w so that $d\sigma^2 = dr^2 - dw^2$ where $w = a\cosh\chi$, $r = a\sinh\chi$ and $w^2 - r^2 = a^2$. This plots as a hyperboloid of revolution with vertex V at $\chi = 0$ and $w = a$. Select an arc-parameter χ_o at a point P_o with coordinates w_o and r_o. Like the spherical case, consider a tangent line at the half-angle $\chi = \chi_o/2$. The slope of the line segment V-P_o is the same as the slope of the curve at a tangent point T at $\chi_o/2$. That is:

$$\frac{dw_t}{dr_t} = \frac{r_t}{w_t} = \frac{\sinh(\chi_o/2)}{\cosh(\chi_o/2)} = \tanh\frac{\chi_o}{2} = \frac{(\cosh\chi_o - 1)}{\sinh\chi_o} = \frac{w_o - a}{r_o} = slope(VP_o) \tag{8}$$

The tangent line has an intercept on the w axis at point B at described by $w = r\tanh(\chi_o/2) + b = rr_t/w_t + b$. At T, $w_t = r_t r_t/w_t + b$, so value $b = ({w_t}^2 - {r_t}^2)/w_t = a^2/w_t$. Note that:

$${r_t}^2 = a^2\sinh^2(\chi_o/2) = a^2(\cosh\chi_o - 1) = (aw_o - a^2)/2$$

A line from the origin, 0, to the point P_o is given by $w = rw_o/r_o$. This line intersects the tangent line at point $P_1 = (r_1, w_1)$. From equation (8), we have $w_o w_t - r_o r_t = a w_t$. At P_1, $w_1 = r_1 r_t/w_t + a^2/w_t = r_1 w_o/r_o$, so,

$$r_1 = \frac{a^2 r_o}{w_o w_t - r_o r_t} = \frac{a r_o}{w_t}, \quad \textit{and} \quad w_1 = \frac{a^2 w_o}{w_o w_t - r_o r_t} = \frac{a w_o}{w_t}$$

Then the Minkowski length of BP_1 is $(\Delta\sigma)^2 = (\Delta r)^2 - (\Delta w)^2 =$

$$\mid BP_1 \mid^2 = \left(\frac{r_o a}{w_t} - 0\right)^2 - \left(\frac{w_o a}{w_t} - \frac{a^2}{w_t}\right)^2 = \frac{a^2}{w_t^2}\left(2aw_o - a^2 - w_o^2 + r_o^2\right) = \frac{2a^2}{w_t^2}(aw_o - a^2) = \frac{4a^2 r_t^2}{w_t^2}$$

Finally,

$$\mid BP_1 \mid = \frac{2ar_t}{w_t} = \frac{2a\sinh(\chi_o/2)}{\cosh(\chi_o/2)} = 2a\tanh\frac{\chi_o}{2} = \rho \tag{9}$$

So, the H^3 isotropic radius ρ is the Minkowski length of the line BP_1.

The hyperbolic plane H^2 can be modeled by hyperboloid of revolution like the Minkowski model above. A more popular, artistic and intuitive way to picture hyperbolic spaces is by way of 'Escher' type graphics in two dimensions where circumferential scale changes with radius according to the isotropic metric factor $\sqrt{A(\rho)} = [1 - (\rho/2a)^2]$. Spherical spaces have the property that the circumference of circles grows less rapidly than $C = 2\pi\rho$, and hyperbolic spaces grow more rapidly than for Euclidean spaces. Repetitive unit figures (like 'Angles and Devils' or 'Circle Limits') become very tiny near the boundary of a unit 2-disk so that effective circumference approaches infinity – angular space grows with radius. [9]. This is not just a toy model or 'art,' the Lobachevsky space L^2 or H^2 is a well developed mathematical arena from the 1800's and early 1900's. The Escher model represents the math of the Poincaré or conformal disk mode for which geodesics are circular arcs that intersect the boundary $S^1 = \partial D^2$ at $90°$ angles. For any geodesic curve, there are an infinite number of other geodesics which do not meet it (there are multiple parallels). Distances between two points are well defined and can also be represented using complex numbers. In the 1800's, other surfaces of negative Gaussian curvature were also discovered such as the pseudosphere or bugle surface. [6]. But they are not true analogues of S^2. The true analog is hyperbolic 2-space H^2 or hyperbolic plane which cannot be embedded in Euclidean space, E^3. [8] . The bugle surface is a tractrix curve rotated about its asymptote, is locally isomorphic to H^2 but only represents a small portion of the whole Poincaré disk area. That is, the 'horocyclic sector' fraction of the Poincaré model is isometric to the pseudosphere. [10].

References

[1] C.W. Misner, K.S. Thorne, and J.A. Wheeler, Gravitation, Freeman, San Francisco, 1973 (pgs. 614, 723-725).
[2] C. Møller, The Theory of Relativity, Clarendon Press, Oxford, 1952.

[3] D. Laugwitz, Differential and Riemannian Geometry, Academic Press, N.Y., 1965 (p. 117).
[4] Hilbert's Theorem: "A complete analytic surface, free from singularities, with constant negative Gaussian curvature, cannot exist in three-dimensional Euclidean space." A proof of Hilbert's Theorem in English can be found in: An Introduction to Differential Geometry, T.J. Willmore, Oxford University Press, 1959.
[5] Pictures of the Euclidean surfaces representing constant negative curvature and the portions of the hyperbolic plane which they map are shown in Felix Klein's book, Vorlesungen *Ü*ber Nicht-Euclidische Geometrie, Verlag von Julius Springer, Berlin 1928, p. 286.
[6] The first mappings of portions of the hyperbolic plane to Euclidean surfaces (e.g., the 'pseudosphere' or 'bugel-surface') were discussed by Ferd. Minding, Journal fur die reine und angewandte Mathematik, 19 (1839) pp 370-387.
[7] David L. Peterson, "Geometrical interpretation of the isotropic radial coordinate" and "Geometry of the Hyperbolic Universe," Dec. 1979, unpublished. email: davepeterson137@gmail.com , Boulder, Colorado.
[8] Barrett O'Neill, Elementary Differential Geometry, Academic Press, N.Y., 1966.
[9] C. Davis, E. Ellers, Editors, "The Coxeter Legacy–Reflections and Projections, American Mathematical Society, Providence, Rhode Island, 2000.
[10] The Pseudosphere, http://math.uncc.edu/ droyster/math3181/notes/hyprgeom/node69.html .

DECOMPOSITION OF THE PERIHELION SHIFT OF GENERAL RELATIVITY

DAVID L. PETERSON

Abstract. The calculation of the perihelion shift of the planet mercury in general relativity theory ('GR') only uses weak field physical concepts below the full power of the theory. In terms of common metric components, gravitational red shift is due to non-linearity in time, light deviation is due half to time and half to space contributions. But perihelion shift uses a blending of space and time which only crudely might be called 2/3rds time and 1/3rd space contributions. This result is not coordinates invariant and depends on the choice of metric form It could be argued that for weak field cases the principle of equivalence combined with special relativity are sufficient to explain red shift, bending of starlight, and even the perihelion precession. Overall confidence in GR depends on additional tests beyond the classical tests.

1. Equivalence:

On November 18, 1915, Albert Einstein performed the first physically correct calculation of the residual 'anomalous' forward precession of the planet mercury that had been previously unexplained using just Newtonian mechanics [1]. Although retrodictive, this was a major accomplishment of the new general theory of relativity. Of the three classical tests of general relativity theory ('GR'), only the perihelion shift depends on approximations to the solutions to the field equations that are usually claimed to go beyond special relativity and the principle of equivalence first-order weak-field contributions. For this calculation, he used an approximate isotropic Cartesian metric form and did not yet need to know the full proper field equations of general relativity. That is, in empty space outside mass sources, the Ricci tensor is $R_{\mu\nu} = 0$. It was a week later when he finally stated the correct general field equations that had eluded him for several years $G_{\mu\nu} = R_{\mu\nu} - g_{\mu\nu}R/2 = -\kappa T_{\mu\nu}$ that include a previously missing 'trace term' or Ricci scalar curvature, $R \equiv R^{\mu}{}_{\mu}$ (and $\kappa = 8\pi G/c^4$). These full equations were then solved by Karl Schwarzschild for an exact exterior and interior stellar solution and were presented on his behalf by Einstein on January 16, 1916. Schwarzschild was in the German army at the Russian front and died several months after his discoveries.

Einstein proposed the key 'Equivalence Principle' for universality of free fall in 1907. Equality of gravitational mass and inertial mass is equivalent to stating that the force experienced in an accelerated frame of reference is equivalent to that locally experienced

Date: 31 January, 2011.

email: davepeterson137@gmail.com. Paper updated to March 24, 2013.

in a gravitational field, and the acceleration of a body in a gravitational field must be independent of its composition. In metric form around a spherical mass source, the principle of equivalence implies that

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = +(1 - \frac{2m}{r})c^2dt^2 - d\ell^2 \tag{1}$$

where $m = MG/c^2$ (and a frequent convention is to set $c \equiv 1$ and $G \equiv 1$). The space term $d\ell^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2d\Omega^2$ where $d\Omega^2 = d\theta^2 + sin^2\theta d\phi^2$ is an element of solid angle. For negligible gravitational potential, the metric becomes the Lorentz metric of special relativistic Minkowski space. With this metric sign convention, the metric distance element could be called $c^2d\tau^2$ or just $d\tau^2$. Notice that division of the metric by dt^2 gives:

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

where Γ could be called a "gravitational Lorentz factor" and u is a velocity accompanying each potential. To low order, one could say:

$$\frac{dt}{d\tau} = \Gamma \simeq (1 + \frac{MG}{c^2r} + \frac{u^2}{2c^2}), \quad or \quad \Delta t \simeq \Delta\tau(1 - \Delta\phi + \Delta(u^2/2c^2)) \tag{3}$$

The metric (1) produces the gravitational red-shift and half of the full light deflection of Einstein GRT. It is commonly known that special relativity ('SR') mechanics contributes a 1/6th factor to the final Einstein result for perihelion shift. Just equivalence g_{oo} by itself in (1) or the α term in equation (16) contributes a 2/3rd factor to perihelion shift. These contributions cannot be added together because the metric equations already contain special relativity as a limiting case. Every small region of space-time is locally Lorentzian. What causes the additional 1/3rd portion? Is it space contraction, finite propagation speed of gravity, energy density in the gravitational field, non-inverse-square field, the principle of general covariance, or something else?

Even though the metric (1) contains Equivalence and SR, it doesn't yet contain them in the right way. It is possible to explain gravitational red shift just using the principle of energy-conservation or the principle of equivalence. Newtonian gravitation itself is due to the effect of the principle of equivalence on time, $g \propto -\nabla h_{oo}$, where the weak field metric $h_{oo} = g_{oo} - \eta_{oo}$, and $\eta_{\mu\nu}$ is the special relativity Lorentz metric. [5]. So, metric (1) has equivalence and SR for its time portion. But, SR also produces space contraction, and that concept analog is missing and needs to be added to the gravitational metric. The bending of starlight and the time delay of radar can be explained using equivalence combined with special relativity (SR) in the 'right way' [4].

2. Decomposition and Non-Contributions:

Some textbooks and papers discuss several basic physical concepts that aid an intuitive physical understanding of weak field general relativity. For example, the mechanics text by Goldstein [3] shows that if a small inverse cube central field were to exist, it could produce a forward perihelion precession of the correct value. This is actually a relevant observation, because the effective general relativistic potential can be cast into a form in which it shows an inverse cube field [8]. However, the potential is also velocity dependent due to conserved orbital angular momentum. Just adding a contribution of special relativistic motion into an otherwise Newtonian inverse square gravitational field can produce one-sixth of the Einstein value of ~ 43 seconds of arc per century for perihelion precession. An early concept from the 1800's was that gravity might not be perfectly inverse square, and only a part in 10^{-7} added to the power 2.0 could achieve the noted perihelion shift. But, it was shown by 1900 from lunar orbit observations that inverse square is much more exact than this [19]. Another possibility noted by Dicke was that solar oblateness can also affect perihelion precession– but with inverse fourth power field contribution [2]. There was strong concern about the strength of this effect until nearly 1974 with some controversy still remaining for another decade after that [19].

In intuitively interpreting and decomposing weak-field effects, a common trap is to assume that the gravitational field is analogous to an electric field and therefore has a local energy density contributing to the gravitational field. That is, in electrostatics, the energy density is $\epsilon E^2/2$. In Newtonian gravitation, $\nabla \cdot \vec{g} = -4\pi G\rho$, and energy density would be $-g^2/8\pi G$. This negative density could be integrated over spherical shells surrounding a mass. So, an initially inverse square field is altered into something more complicated by successive shells of negative energy density acting as a source for a given radius ("mass-feedback"). This approach doesn't work and doesn't help explain perihelion shift. Local gravitational energy cannot be a relativistic source term for the effective gravitational field. The reason is obvious: by the principle of equivalence, there is always a frame of reference from which any local gravitational field can be made to vanish. No local field means no local energy density [2]. A particle in free fall (like the planet Mercury) is in an an accelerated frame which in all parts of its orbit transform away the g field and hence g^2. And, even in the electrostatic case, there is ambiguity about where the energy is stored [13]. Energy can be considered as $U = \int \rho\phi dV/2$ [14] , which is mathematically equivalent to $U = \epsilon_o \int E^2 dV/2$. But these forms have different interpretations: energy is where the charges are versus where the field is. There are still occasional published articles based on the concept of mass-feedback – sometimes under the name "pseudo-Newtonian" gravitation [16]. It is not equivalent to GRT and gives different answers to problems.

The suggestion that a contribution may come from the finite speed of propagation of gravitation [9] was discussed in the 1800's but most notably in 1898 by Paul Gruber [11] who obtained the correct degree of perihelion shift prior to Einstein. His work was flawed with conclusions that did not follow from his premises, and his orbit equation for $u'' + u = N(u)$ was different from that of GR. So far, no one has been able to clearly polish his theory so that it is really valid and also consistent with the bending of starlight. There is inference that the speed of gravitation should be c, but there is still no direct measurement that this is true. It is an unproven contribution. It is possible that "mass-feedback" and finite speed of propagation are already built into GR so that they cannot be discussed separately.

Einstein's "Principle of General Covariance" says that fundamental physical laws should not depend on particular coordinate systems and should be invariant under differentiable coordinate transformations. The principle of relativity can be extended for accelerated motions. Although this principle motivated Einstein's development of GR and is presented in most texts on the subject, it is not universally accepted as a valid principle [15]. It contains no real physical content and is not even a symmetry principle. It also doesn't lead directly to GR– Newtonian gravity can be cast into a generally covariant form, and Lagrangian mechanics has always been invariant under arbitrary spatial transformations. A search for a general covariance contribution to perhelion shift would be in vain. What has to be added is that the laws of physics should have a tensor form involving 4-dimensional pseudo-Riemannian spacetimes. In particular, a final axiom would be that the metric has to be a solution of Einstein's field equation, $G = 8\pi T$. The Einstein tensor $G_{\mu\nu}$ would change its expression under coordinate transformations, but its vanishing would still be preserved in empty space. From today's perspective, perhaps what Einstein would now say is that invariance under the Lorentz group for special relativity could be broadened to a general group for gravitation theory. But such power is not required for weak field perihelion shifts nor for light bendings. As opposed to Einstein relativity, the Lorentz-Poincare version has preferred reference frames (called the "aether"). Also, in cosmology, one could speak of a preferred reference frame– the one at rest with respect to uniform incidence of cosmic black-body radiation. Some versions of quantum mechanics like preferred reference frames. But the majority in the physics community would prefer general covariance. This argument is relevant here because the Schwarzschild coordinates give a different space versus time decomposition of the perihelion shift of Mercury. Is such a decomposition meaningful?

3. From Special Relativity:

A teaching journal article in 1960 by Schiff [4] showed how to derive first order space and time dilations due only to relativity and the principle of equivalence and applied it to the correct bending of starlight. With the gravitational field replaced by an accelerating frame of reference, essentially the Lorentz contraction in special relativity corresponds to observed radial length contraction in a gravitational field, and time dilation corresponds to time expansion in a gravitational field [6]. Lengths perpendicular to the gravitational field are left unchanged. The bending of starlight is due to equal contributions of distortions in time and space. Einstein's early papers only considered time dilation due to the principle of equivalence by itself. The double contribution for bending of starlight was discovered later and appended to the general theory of relativity in time for the proper verification begun by Eddington using a solar eclipse and then later refined by others.

General Relativity is built upon special relativity. The postulates of special relativity can be selected in a variety of ways. Einstein's 1905 postulates were the principle of relativity ('PR'– equivalence of inertial frames) and invariance of the speed of light, c. One could also begin with the 'Lorentz-Fitzgerald Contraction' and PR [10], or with Maxwell's equations and Lorentz Transformations ('LTs') and PR. Or, start with the Lorentz metric for an increment of proper time, τ (which itself is determined in a 'rest-frame' with no space separation between events close in time). The usual textbook approach is from Einstein postulates for light-like intervals to the Lorentz Transformations and from there to length contraction and time dilation and other effects. The invariance of the speed of light is automatically built into the Lorentz transformations. That is, suppose 'x' is a space coordinate in the same direction as a moving frame with velocity v. Then, the speed of light is observed to be:

$$c' = \frac{dx'}{dt'} = \frac{\gamma(dx - vdt)}{\gamma(dt - vdx/c^2)} = \frac{(dx/dt - v)}{1 - \frac{vdx/dt}{c^2}} = \frac{c(c-v)}{(c-v)} = c. \tag{4}$$

More trivially, if the SR metric is $(cd\tau)^2 = (cdt)^2 - d\ell^2$ and light-like motion has $d\tau = 0$, then $d\ell/dt = c$. This invariance is broken in GR where a distant observe would see a reduced c by a gravitational field. This in turn would enable some GR calculations to be performed using an effective index of refraction, n. In special relativity, an initial light-like metric with $d\tau = 0$ allows lengths to be measured using light rays with a constant speed of light, $\Delta x = c\Delta t$. Time dilation could be found directly from the Lorentz metric.

$$c^2 d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu = +c^2 dt^2 - dx^2 \quad \Rightarrow \left(\frac{d\tau}{dt}\right)^2 = 1 - \left(\frac{v}{c}\right)^2 = \frac{1}{\gamma^2} \quad \Rightarrow dt = \gamma d\tau \tag{5}$$

Perhaps the simplest example of SR length contraction 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. $c^2\Delta t^2 - \Delta x^2 = c^2\Delta t'^2 - \Delta x'^2$. Since $\Delta x = 0$ in S, the Δt is proper time $= \Delta\tau = L/v$. $L' = v\Delta t'$ and $\Delta x' = L'$. Then:

$$(cL/v)^2 - 0 = (cL'/v)^2 - L'^2, \quad L^2 = L'^2(1 - v^2/c^2), \quad L = L'/\gamma. \tag{6}$$

Again, this is consistent with time dilation:

$$\Delta\tau = \frac{L}{v} = \frac{L'}{\gamma v} = \frac{v\Delta t'}{v\gamma}, \quad \Delta t' = \gamma\Delta\tau. \tag{7}$$

The 'Principle of General Relativity' says that a local inertial system experiencing a constant gravitational force is equivalent to a noninertial system undergoing constant acceleration (relative to the fixed stars). 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.

Technically, it is not true that free fall in a gravitational field is the same as the effects of an observer's acceleration [15]. 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 as discussed above, 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 [10].

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 [4]. 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 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:

$$T_B \simeq T_A[1 + (v_B^2 - v_A^2)/2c^2] \simeq T_A(1 + gh/c^2) \simeq T_A[1 + GM/c^2 r_B - GM/c^2 r_A] \tag{8}$$

This period elongation, $T_B > T_A$, is called 'Red Shift.' This concept has been proven to apply to both light and to 'matter waves' as well [17]. Phase difference measurements in an atom or neutron interferometer are the same as those accumulated using conventional clocks following the same paths. 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,

$$L_B \simeq L_A[1 - (v_B^2 - v_A^2)/2c^2] \simeq L_A(1 - gh/c^2) \simeq L_A[1 - GM/c^2 r_B + GM/c^2 r_A] \tag{9}$$

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

$$\frac{dt'}{d\tau} \simeq \frac{T_B}{T_A} \simeq [1 + GM/c^2 r] \quad and \quad \frac{dr'}{dr} \simeq \frac{L_B}{L_A} \simeq [1 - GM/c^2 r]. \tag{10}$$

These can be assembled by components into a metric:

$$\mathbf{d\tau^2 \simeq dt'^2(1 - 2m/r) - dr'^2(1 + 2m/r) - dr'_\perp{}^2} \tag{11}$$

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 (11) 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 [6]:

$$\frac{\nu_A - \nu_B}{\nu} = \frac{T_B - T_A}{T} = \frac{\lambda_B - \lambda_A}{\lambda} = \frac{\phi_B - \phi_A}{c^2} \tag{12}$$

Duration is a number of periods and length is a number of wavelengths. So this result is consistent with the first order length transformation (9). 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,

$$\frac{dr'}{dt'} \simeq \frac{dr}{d\tau}\left(\frac{1 - m/r}{1 + m/r}\right) \quad \Rightarrow c' \simeq c(1 - 2m/r) \tag{13}$$

it slows down in near field. Light speed c is a local constant, but at distance separation it is non-constant and non-isotropic.

4. Calculations and 'PPN':

The exact exterior solution $g_{\mu\nu}$ of Einstein's equations for general relativity theory for a non-rotating spherical central mass is given by the Schwarzschild metric

$$d\tau^2 = +(1 - \frac{2m}{r})c^2dt^2 + \frac{dr^2}{(1 - \frac{2m}{r})} - r^2d\theta^2 - r^2\sin^2\theta d\phi^2 \tag{14}$$

The metric (14) can be transformed into a more Euclidean spatial form through change of variable for the radial coordinate, $r = \rho(1 + m/2\rho)^2$. This is a differentiable coordinate transformation and gives:

$$d\tau^2 = \frac{(1 - \frac{m}{2\rho})^2}{(1 + \frac{m}{2\rho})^2}\, c^2dt^2 - (1 + \frac{m}{2\rho})^4\, d\ell^2 \tag{15}$$

This metric *eqn.*(15) is called the "isotropic" metric or the "isotropic Schwarzschild" line element to distinguish it from other isotropic metrics such as that used for cosmology. An isotropic metric form has the virtue that the relativistic local curved space coordinates look more like Euclidean classical coordinates x, y, z, or r θ, ϕ. There is a general preference to use isotropic coordinates when discussing solar system astronomy.

Notice that the usual transformation $r = \rho(1 + m/2\rho^2)$ can be approximated to first order simply as $r' \simeq \rho + m$ and $dr' \simeq d\rho$. This then yields the isotropic PPN approximation (11) with parameter values $\alpha = \beta = \gamma = 1$. For the sun with mass 1.99×10^{30} kg, gravitational radius is a relatively tiny distance $m = MG/c^2 = 1480$ meters. The eccentricity of the orbit of mercury is high at $e = 0.206$, and the mean radius of orbit is 57.9 Mm. So the max minus min planetary excursion is two elliptical focii distances $\simeq 23$ Mm which is substantially larger than the gravitational radius offset correction. So essentially $r' \simeq \rho \simeq r$ with no real need to differentiate the types of radius.

An approximation to the isotropic line element 'to post-Newtonian accuracy' is given by:

$$d\tau^2 = \left[1 - 2[\alpha]\left(\frac{m}{\rho}\right) + 2\beta\left(\frac{m}{\rho}\right)^2\right] dt^2 - \left[1 + 2\gamma\left(\frac{m}{\rho}\right)\right] d\ell^2 \tag{16}$$

In Einstein's GR with this metric form, the values of α, β and γ are unity. β and γ are two of the ten parameterized post Newtonian or 'PPN' parameters and are are described as the degrees of 'nolinearity in the superposition law for gravity,' and 'space curvature' per unit test mass. The use of a diagonal and isotropic equations is referred to as the 'standard PN gauge.' Note that the parameter β is appropriate in isotropic coordinates but would not naturally appear in coordinates for time in equation (14). $\alpha = 1$ is usually not mentioned at all but serves a purpose of tagging the time versus space contributions in calculations of solar system effects in the isotropic metric view [2]. This contribution is considered trivial and expresses the principle of equivalence (1). The relative contributions of astronomical effects due to dilation of time versus that of space does depend on the form and coordinates chosen for the metric [7] – for example on equation (14) versus (16). The gravitational red-shift of electromagnetic radiation depends only on the Newtonian potential and conservation of energy– or on just the parameter α. Every metric theory of gravity automatically predicts the same degree of red-shifting because all of them are based on the principle of equivalence. The bending of starlight depends only on the Newtonian gravitational potential and the PPN parameter γ or on the combination $(\alpha + \gamma)$. The perihelion shift depends on a more complex combination $\propto (2\alpha - \beta + 2\gamma)$ resulting in $6\pi m/a \sim 43$ seconds of arc per century, where a is the mean radius of the orbit of mercury. This PPN Isotropic result might be interpreted as having 2/3rds contribution to perihelion shift from space and a net 1/3rd from time. But would this mean anything? In terms of the concept of general covariance, might the Schwarzschild coordinate PPN result be different? Yes, it is. Is one view more correct than the other?

Some texts on GR prefer to calculate the perihelion precession using the exact Schwarzschild metric instead of the isotropic metric [5]. In outline, the calculation is as follows. The motion of a planetary body is in free-fall and therefore follows the world line or geodesic of the curved spacetime metric. For simplicity, let $\theta = \pi/2$ on the equatorial plane. The orbit is given by the Euler-Lagrange equations for $\delta \int ds = 0$. Or, with more ease, Hamilton's principle can be used as a general definition of geodesics– an equivalent variation formula: $\delta \int L(x_i, \dot{x}_i, t)dt = 0$. The integrand of this integral is $g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$ which has the appearance of just showing kinetic energy (where $\dot{x} = dx/d\tau \Rightarrow dx/d(c\,\tau)$, and τ is proper time) [5]

[9].

$$d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu \Rightarrow \quad g_{oo}\dot{t}^2 - g_{rr}\dot{r}^2 - r^2\dot{\phi}^2 = \dot{\tau}^2 = 1 \tag{17}$$

[For light rays, null geodesics obey $d\tau^2 = 0$, and the right hand value is zero].

The Euler-Lagrange equations yield conservation of angular momentum/unit mass, $h = r^2\dot{\phi} = constant$, and $g_{oo}\dot{t} = \eta = cnst$. This means that $\dot{t} = \eta/g_{oo}$ so that the next step is to multiply through by g_{oo}. Then the η term will be isolated so that its derivative is zero. The resulting Kepler type orbital differential equation with variable $u = 1/r$ and $u' = du/d\phi$ can then be shown to be: $\mathbf{u'' + u = m/h^2 + 3mu^2}$.

For classical Kepler motion, $u'' + u = GM/H^2$ where $H = r^2 d\phi/dt = cr^2 d\phi/(dct) = ch$ is mass-normalized angular momentum. In polar coordinates, the usual form for an ellipse is $r = ed/(1 + e\cos\phi)$ where d is directrix, and eccentricity is $e = c/a$ is focus/major vertex distance. In the solar system, $r = H^2/GM[1 + e\cos\phi]$ or $u = MG[1 + e\cos\phi]/H^2$. Since $m = MG/c^2$ is a weak perturbation of the order of 10^{-7}, terms with m^2 or higher are often negligible in solar system physics. As will be shown later, after some calculation from the orbital equation with perturbation $3mu^2$, the perihelion shift per revolution, δ, is obtained resulting in key terms like $(m/h)^2$. Use $H^2 = GMa(1 - \epsilon^2)$ to get:

$$\delta = (2\pi)3\frac{m^2}{h^2} = 6\pi\frac{MG\,MG}{c^2\;\,H^2} = 6\pi\frac{MG}{c^2 a(1-\epsilon^2)} = 6\pi\frac{m}{a(1-\epsilon^2)}. \tag{18}$$

For light rays, the equation would be $u'' + u = 3mu^2$.

It is a convention to discuss PPN using the isotropic metric form, but at least one text uses a Schwarzschild PPN coordinate metric instead [10]. The α parameter is the same for both types of coordinate systems, but the meanings of β and γ are different; and the Einstein values here are $\beta = 0$ and $\gamma = 1$. The PPN results appear differently, e.g., perihelion advance $\propto (2\alpha^2 + \alpha\gamma - \beta)/\alpha$, and light bending $\propto 2(\alpha + \gamma)$. Because the $(\alpha + \gamma)$ factor appears using both isotropic and Schwarzschild coordinates, it makes sense to say that the bending of starlight is half due to time and half due to space curvature. But the contribution from space curvature to the perihelion shift isn't as clear. For just the principle of equivalence metric (1), the perihelion precession is $\delta = 4\pi/a(1 - \epsilon^2)$ which is 2/3-rds the Einstein value. For first order "linearized" GRT ($\alpha = 1$ and $\gamma = 1$ but no β term), the precession is 4/3rds the Einstein value for isotropic coordinates. But, as shown below, the **correct** 3/3rds full value can be obtained with no beta term using Schwarzschild coordinates!

The metric forms essentially depend on terms resembling gravitational potentials which could be labeled as $U(\rho) = -MG/\rho c^2$ for isotropic coordinates and $V(r) = -MG/rc^2$ for Schwarzschild coordinates. The Schwarzschild radius is a great circumference divided by 2π and is generally greater than the isotropic radius in value. It could be argued that the first order Schwarzschild metric in Schwarzschild coordinates provides the proper physical understanding for the weak-field perihelion shift. Transforming to isotropic coordinates would then introduce a new $\beta_{iso} \simeq (\beta_{schw} + \gamma_{schw}\alpha)$ which would also lead to the correct isotropic perihelion shift. This physical intuition could be based on analogy to special relativity and the principle of equivalence using logic similar to that of Schiff [4]. The use of Schwarzschild coordinates would be more appropriate because it is the radial coordinate that experiences distance contraction while angular distances remain unchanged.

The testing of the broad concepts and results of general relativity then depend on tests beyond classical weak-field GR including: the Lense-Thirring "gravito-magnetic" effect, strong field binary pulsars (such as PSR 1924+16), gravitational radiation, LIGO, black-holes, higher order PPN, post Keplerian parameters ["PK," such as orbital decay rate], gravitational lenses, and strong field Shapiro delay. For the new pulsar PSR J0737-3039, perihelion shift is $\sim 17°/year$ [12]. The Hulse-Taylor pulsar had $4.2°/year$.

5. Perihelion Shift Decomposition Calculation:

Consider a Schwarzschild type Metric with Schwarzschild coordinates of the form (14) but instead using just the first order linear approximation to g_{rr} as in equation (11) so that the new

$$\mathbf{g_{rr} = (1 + 2\gamma m/r), \text{ and } \quad g_{oo} = (1 - 2\alpha m/r)}$$

with the Greek letters again mainly serving as coefficient labels for the space and the time contributions separately. Processing this metric the usual way (as above) gives more complex equations than just using the exact Schwarzschild metric without labels. Begin with the modified equation (17) which now will become:

$$\left(1 - \frac{2\alpha m}{r}\right) = c^2\eta^2 - [(g_{oo})(g_{rr})]r'^2\,\frac{h^2}{r^4} - \frac{h^2}{r^2}\left(1 - \frac{2\alpha m}{r}\right) \tag{19}$$

where,

$$(g_{oo})(g_{rr}) = (1 + 2\gamma m/r - 2\alpha m/r - 4\alpha\gamma m^2/r^2) = (1 + 2\gamma mu - 2\alpha mu - 4\alpha\gamma m^2u^2)$$

Express the equations in terms of $u = 1/r$ and differentiate $u' = \partial u/\partial\phi$ and re-arrange to obtain:

$$u''\,[(g_{oo}g_{rr})(u)] + u = \frac{\alpha m}{h^2} + 3m\alpha u^2 - (mu'^2)(\gamma - \alpha - 4\alpha\gamma mu) \tag{20}$$

For convenience of orbital calculations, the terms at left have to be of the form $u'' + u$. To simplify, label $\zeta = 2(\gamma - \alpha)$, then:

$$u'' + u = N(u) = \frac{\alpha m}{h^2} + 3m\alpha u^2 - \zeta mu''u - (mu'^2)(\gamma - \alpha - 4\alpha\gamma mu) + ...$$

Each of the right hand side perturbations to a pure elliptical orbit can be evaluated separately using the method of successive approximations. That is, first use a trial basis solution of $u_o = (1 + e\cos\phi)/a$ where $a = H^2/(MG\alpha) = h^2/m\alpha$. Substituting into $u'' + u$ automatically yields the first term on the right side, $\alpha m/h^2$ which produces a standard ellipse when $\alpha = 1$. But also evaluate u_o in each isolated perturbation term on the right to be used for comparison later. Then form a second perturbed approximation (label 'p') expression and evaluate it for $u_p'' + u_p$ where (for example),

$$u_p = A + B\phi \sin\phi + C\cos(2\phi)$$

$$u_p'' + u_p = A + 2B\cos\phi - 3C\cos 2\phi$$

The unknown coefficients A, B, and C can then be evaluated by term-by-term comparison [18] with the middle term of u_p being the key term causing precession of the elliptical orbit. The B term is then found to be:

$$B = \frac{m^3 e\alpha^2}{h^4}(3\alpha + \gamma - \alpha = 2\alpha + \gamma).$$

The "3α" term represents the standard Einstein factor. The final contribution for the $\phi\sin\phi$ term will be some tiny number – call it δ. Then consider the ellipse term:

$$1 + e\cos(\phi - \delta\phi) = 1 + e[\cos\phi\cos(\delta\phi) + \sin\phi\sin(\delta\phi)] \simeq 1 + e\cos\phi + \delta\phi\sin\phi$$

$$u \simeq u_o + u_p = \frac{m\alpha}{h^2} + A + \frac{me\alpha}{h^2}(\cos\phi + \frac{m^2\alpha(2\alpha+\gamma)\phi\sin\phi}{h^2}) + C\cos 2\phi$$

The perihelion shift per revolution is then $\simeq 2\pi\delta$ where $\delta = (\mathbf{m}/\mathbf{h})^{\mathbf{2}}\alpha(\mathbf{2}\alpha + \gamma)$. This result can be further processed using equation (18). This yields the correct Einstein perihelion precession value and has the same decomposition as suggested by Robertson and Noonan [10]. It suggests an approximation of contributions to the perihelion shift as 2/3rds time and 1/3rd space effects.

When this method is used for the mu'^2 term, it turns out to give B = 0 and therefore doesn't contribute to precession. The other terms are order m^2 or higher and are too weak to contribute. The perturbation terms like mu^2 contribute +1/3rd the Einstein value while terms like $mu''u$ contribute -1/6th the Einstein precession each.
One important observation is that the decomposition of the perihelion shift is not coordinate invariant. If the Isotropic PPN metric is used, then the result is essentially 2/3rds due to space and 1/3 due to time– a different result. How one looks determines what one gets. Traditional general relativity would say that there is no unique decomposition into space and time contributions for perihelion shift because the concept violates the principle of general covariance.

Alternative Approach:

Instead of keeping the perturbation terms with u", the u" term could instead be isolated by multipying the equation (20).by the reciprocal:

$$[(g_{oo}g_{rr}]^{-1} = (1 - \zeta mu + (4\alpha\gamma + \zeta)m^2u^2), \text{to obtain:}$$

$$u'' + u = N(u) = \frac{\alpha m}{h^2} + 3m\alpha u^2 + \zeta m u^2 - \zeta m u'^2/2 - \zeta\alpha\frac{m^2 u}{h^2} + ... \tag{21}$$

There are higher terms which will be negligible, but the last term isn't one of them. Note that if $\gamma = \alpha$ as in GRT, then $\zeta = 0$, and most of these terms vanish leaving just the first two– the standard terms of GRT. There are several terms in the key form mu^2, and they can be combined together. There are also several terms in the resulting shift which involve $(m/h)^2$, and they can be combined together. The result for perihelion shift per revolution is:

$$\delta = \frac{m^2}{h^2}\left[3\alpha^2 + \zeta\alpha - \frac{\zeta\alpha}{2}\right] = \frac{\alpha m^2}{h^2}[2\alpha + \gamma] \tag{22}$$

This is the same result as in the previous calculation and would suggest a decomposition of 2/3rds time and 1/3rd space contribution for the perihelion shift. For $\alpha = \gamma = 1$, this gives the perturbation $3mu^2$ and the correct GRT precession value.

Note that Schiff himself stated that that perihelion motion required a $\beta(m/r)^2$ contribution– but that would only be true for isotropic PPN form and not for Schwarzschild coordinates where radius is treated differently from angles. He also said that "an equation motion for a particle of finite rest mass" is needed for sub-light speeds [4].

References

[1] Abraham Pais, Subtle is the Lord, The Science and Life of Albert Einstein, Oxford University Press, N.Y., 1982.

[2] C.W. Misner, K.S. Thorne, and J.A. Wheeler, Gravitation, Freeman, San Francisco, 1973 (pgs. 184, 435, 467, 607, 840, 1115).

[3] Herbert Goldstein, Classical Mechanics, Addison-Wesley Publishing Company, Reading Massachusetts, 1950 (pgs. 91 and 214).

[4] L. I. Schiff, "On Experimental Tests of the General Theory of Relativity," American Journal of Physics Vol. 28 No. 4, 1960, p 340.

[5] Ronald Adler, Maurice Bazin, Menahem Schiffer, Introduction to General Relativity, McGraw-Hill, N.Y., 1965, pgs. 56, 124.

[6] David L. Peterson, "General Relativity for Pedestrians," 1974,unpublished. Portions also found in "General Physics 201-4" CU Lifelong Learning, Independent Study Syllabus 1977.

[7] David L. Peterson, An Elementary Introduction to the General Theory of Relativity, book, 1977–unpublished. Additional sections to up to 2011.

[8] Robert M. Wald, General Relativity, The University of Chicago Press, 1984 (pg. 139).

[9] Jerry B. Marion, Stephen T. Thornton, Classical Dynamics of Particles and Systems, Fourth Ed., Saunders College Publishing, Harcourt Brace & Company, 1995. pp. 235.

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

[11] Paul Gruber, "The propagation of gravity in space and time," Zeitschrift fur Mathmatik und Physik, 43, pp. 93-104, 1898.

[12] M. Kramer, et. al., Tests of General Relativity from Timing the Double Pulsar Science 6 October 2006: Vol. 314 no. 579. pp. 97-102

[13] P. C. Peters, "Where is the energy stored in a gravitational field?" American Journal of Physics, 49(6), 564-569 (1981).

[14] John Reitz, Frederick Milford, Foundations of Electromagnetic Theory. Addison-Wesley Publishing Company, Inc. 1960.
[15] 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].
[16] Michael R.R. Good, Extremal Black Hole in a Nonlinear Newtonian Theory of Gravity, 11/12/2008. http://arxiv.org/pdf/0811.1805
[17] 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
[18] Chris Pollock, Mercury's Perihelion, March 31, 2003, http://www.math.toronto.edu/ ˜ colliand/426_03/Papers03/C_Pollock.pdf
[19] Clifford M. Will, Was Einstein Right? Putting General Relativity to the Test, Basic Books, N.Y., 1986 (p.93, 102).
[20] M.A. Abramowicz, et. al., "The perihelion of Mercury advance and the light bending calculated in (enhanced) Newtons theory," arXiv:1303.5453 [gr-qc] 21 Mar 2013.

6. Additional Contributions

Abramowicz and Ellis [20] have an interesting approach to the perihelion shift using standard Newtonian gravitation on a curved 3-D geometry. They claim that careful thought might have enabled Gauss to think about perihelion advance long before Einstein. It is sensible to express the Poisson equation for Newton gravity as:

$$ {}^{3}g^{ik}\nabla_i\nabla_k\Phi = -4\pi G\rho, \quad F_i = ma_i \tag{23}$$

for arbitrary 3-geometry. Rather than using just one radial coordinate, r, for curved space it is necessary to consider three possibilities: $r_\star$ for 'geodesic radius,' $\tilde{r}$ for circumferential radius, and $\mathcal{R}$ for curvature radius (in the sense of the Frenet formula). One can deduce these physically from careful measurement of centrifugal acceleration, $a_c = V^2/\mathcal{R}$, and gravitational acceleration, $g = GM/\tilde{r}^2$. In Euclidean space, these radii are the same, but they split in curved space. Analysis leads to a metric based on $ds^2 = dr_\star^2 + \tilde{r}^2 d\phi^2$ which for 3-space becomes:

$$ds^2 = \left(\frac{r-M}{r-3M}\right)^2 dr^2 + r^2\left(1-\frac{2M}{r}\right)(d\theta^2 + \sin^2\theta d\phi^2) \tag{24}$$

Processing this metric leads to the same value for the perihelion advance of Mercury as in Einstein's theory and also the same deflection of starlight as in Einsteins GRT. They mention that Gauss was a master of orbit calculations (Ceres orbit at age 23), actually attempted to measure curvature of space, and derived Bolyai's results first (but didn't publish). If he had thought in this new direction, he could have anticipated perihelion shift.

THE RADIUS OF THE UNIVERSE– A BRIEF SUMMARY

DAVID L. PETERSON

Abstract. The concept of the radius of a spherically symmetric universe can be introduced in several elementary ways. For expanding universes, the radius changes with time. A proper radius lies beyond direct observation and has to be calculated by including the history of stretching space.

1. Newtonian Gravitational Field:

An introduction to basic cosmology can begin with a Newtonian perspective. One old idea in the history of physics is that a star may become massive enough that even light might become unable to escape from its surface. It simply applies the concept of escape velocity from the gravitational field of a massive heavenly body to light rays. This idea originated with John Michell (Professor of Geology) in a letter from 1783. Suppose first that an object of mass m is shot upwards away from a larger body of mass M. The threshold of escape occurs when:

$$E = 0 = KE + PE = \frac{1}{2}mv^2 - \frac{mMG}{R}, \quad so \quad R = \frac{2MG}{v^2} \tag{1}$$

The ultimate escape velocity is the speed of light, v = c, and the radius from which no escape can occur is an early version of "dark star" or black hole but could also be applied to a Newtonian universe as a whole. Consider a diffuse spherical density ρ out to that radius R which would then contain a mass M = $4\pi R^3\rho/3$ where $R = 2MG/c^2$ (which has the appearance of the Schwarzschild "singularity radius"). Therefore,

$$R^2 = \frac{3c^2}{8\pi\rho G} \tag{2}$$

Although easy to grasp, this physics is recognized to be inappropriate and fails to include special or general relativity. But, if one used the currently accepted universal density, $\rho_c \simeq 9.3 \times 10^{-27} kg/m^3$, then the radius would be a plausible 1.3×10^{26} meters $\simeq$ 14 billion light years.

Now suppose that the mass tossed outward is really a spherical shell of matter and let total energy be conserved so that $KE + PE = E_o$. Let $v = HR$ where H is like a "Hubble parameter." Then, per unit mass of shell,

$$\frac{v^2}{2} - \frac{MG}{R} = \frac{H^2R^2}{2} - \frac{4\pi\rho GR^2}{3} = {E_o}' \tag{3}$$

Date: Feb 27, 2011.

email: davepeterson137@gmail.com. Paper updated to April 21, 2011, March 24, 2013.

The 'Gaussian curvature' of a spherical surface is given by $K = 1/R^2$ or $KR^2 = +1 = k$. So, with some advanced knowledge, let the total energy term be replaced by $kc^2/2$ [13]. Then,

$$H^2 - \frac{kc^2}{R^2} = \frac{8\pi G\rho}{3} \tag{4}$$

This equation is very similar to the Einstein general relativity field equation (11), which might aid in its intuitive interpretation.

Another point of interest is that the classical Newtonian gravity field might appear to possess negative energy density. That is, $\nabla \cdot \vec{g} = -4\pi G\rho$, and energy density would be $-g^2/8\pi G$. This negative density could be integrated over spherical shells surrounding a mass with radius a and also the interior field inside. Then,

$$E = -\frac{1}{8\pi G}\left[\int_a^\infty \frac{(MG)^2}{(r^2)^2} + \int_o^a \frac{(MGr)^2}{(a^3)^2}\right] 4\pi r^2 dr = -\frac{3M^2G}{5a} \tag{5}$$

Now, at the Schwarzschild radius, $E = Mc^2 = 2M^2G/R$. So, if a is set equal to R, the negative energy is $\simeq 30\%$ of the positive mass energy of this Newtonian universe – a substantial proportion. It is sometimes claimed that the real universe has a near balance between negative gravitational energy and positive mass energy so that the net total mass-energy is zero. It is possible to have creation from nothing and preserve the "nothing" (e.g., E. Tyron, 1973). In reality, universal energy is very hard to define in general relativity and may be meaningless. [For further discussion on the energy of the universe, see Appendix at end].

Special Case: Suppose the universe consists of pure matter and has a critical rate of expansion for which total energy is $E'_o = 0$. At the present time, let the universal radius $= R_o$ with expansion rate v_o. Energy is always conserved so $v^2/2 - GM/R = v_o^2/2 - GM/R_o = E'_o = 0$. Use $M = 4\pi\rho R^3/3 = 4\pi\rho_o {R_o}^3/3$ and divide by ${v_o}^2$ to get:

$$\left(\frac{v}{v_o}\right)^2 - \frac{8\pi GR^2\rho}{3{v_o}^2} = 1 - \frac{8\pi G{R_o}^2\rho_o}{3{v_o}^2}$$

Introduce a "scale factor" for the size of the universe, $S = R/R_o$ [15]. And let critical density for borderline universal expansion at present be $\rho_{o,c} = 3H_o^2/8\pi G$ where present Hubble $H_o = v_o/R_o$. Then KE + PE = Total energy becomes:

$$\left(\frac{v}{v_o}\right)^2 - \frac{S^2\rho}{\rho_{o,c}} = 1 - \frac{\rho_o}{\rho_{o,c}} \tag{6}$$

This is a special form of the Friedmann equation of general relativity. The last term for ratio of present densities is also called Ω. Notice that the rate of scale $dS/dt = dR/R_o dt = v/R_o = (v/v_o)(v_o/R_o) = H_o v/v_o$, then $v(t)/v_o = t_H dS(t)/dt$ where extrapolated "Hubble Time" is $t_H = 1/H_o$.

How can one find the real universal time from the Big Bang to now, t_o, for this matter-only universe?

$$\frac{dS}{dt} = \frac{v}{R_o} = \frac{1}{R_o}\sqrt{\frac{2MG}{R}} = \sqrt{\frac{2MG}{S{R_o}^3}},\ \ S^{1/2}dS = \frac{2}{3}dS^{3/2} = \sqrt{\frac{2MG}{{R_o}^3}}dt$$

Integrate to get:

$$S^{3/2}(t) = \frac{3}{2R_o}\sqrt{\frac{2MG}{R_o}} = \frac{3v_o t}{2R_o} = \frac{3t}{2t_H}$$

For the present time, now, $S = R/R_o = R_o/R_o = 1$, so

$$S(t) = \frac{R}{R_o} = \left(\frac{3t}{2t_H}\right)^{2/3},\ \ {S_o}^{3/2} = \frac{3t_o}{2t_H}, \quad t_o = \frac{2t_H}{3} \tag{7}$$

Again, this Newtonian result is the same as that obtained by General Relativity. As one approaches the origin of the Big Bang at t = 0, the velocity of expansion becomes infinite. Why should these classical arguments work? The answer is a theorem of George Birkoff in 1923 which states, "the geometry of any spherically symmetric vacuum region of spacetime is a piece of the Schwarzschild geometry" [1]. Or in simpler language, "In an isotropic Universe you can carve out a sphere and ignore the surroundings" [15]. So one particular ball of Newtonian matter is similar to any ball in a huge general relativistic universe.

The concept of initial universal inflation can also be aided by a Newtonian analogy. Consider a cylindrical hole passing all the way through the center of a spherical body in the vacuum of space and drop a rock into the hole. The rock will accelerate as $g = -MG/r^2 = -4\pi\rho Gr/3 = \ddot{r}$. The solution to this equation is just simple harmonic motion like a mass on a spring with angular frequency $\omega = \sqrt{4\pi\rho G/3}$. The rock will pass through the center to the other side and then back again to the starting point to repeat the cycle. Now suppose that instead of the usual gravity due to the mass inside a Gaussian volume, we instead have "anti-gravity." This change in sign will change the dynamics from SHM to exponential motion away from the core. In inflation theory, there is initially a momentary region of "false vacuum" with very high but constant unchanging density, ρ and a negative pressure $p = -\rho$. Einstein cosmology says that a universal radius obeys:

$$\ddot{R} = \frac{-4\pi G(\rho + 3p)R}{3} = \frac{+4\pi G(2\rho)R}{3} \Rightarrow \quad R(t) = ke^{t\sqrt{8\pi\rho G/3}} \tag{8}$$

So, expansion of a constantly growing Gaussian volume is exponential. A grand unified time scale might be $10^{-38}s$ over perhaps 60 e-foldings of expansion.

Alternatively, from equation (6) consider the case that the whole universe is nothing but constant vacuum energy at critical density. Then the right side of the equation is zero and the left side leaves: $v/v_o = S = R/R_o$ or $dR/R = dS/S = H_o dt$ with obvious solution $R = R_o e^{t/t_H}$. Because these deSitter type solutions are driven effectively by anti-gravity, the universe falls down as it expands.

2. Radius of Curvature matches Radius of Gravitational Source:

If a body is tossed into the air from the surface of the earth, then its altitude is given by: $y = y_o + vt - gt^2/2$. Let $x = ct$ so that $y = y_o + v_x x/c - gx^2/2c^2$. Then $dy/dt = y' = v_x/c - gx/c^2 \ll 1$ and $y'' = -g/c^2$. From elementary calculus, the 'radius of curvature' of a function, $y = f(x)$ is given by:

$$R = \frac{[1+(y')^2]^{3/2}}{|y''|} \simeq \frac{1+\epsilon}{g/c^2} \simeq \frac{c^2}{g} \tag{9}$$

For earth gravity, this radius is about one light year; and all parabolic trajectories are really viewed as portions of great circles having this radius. In general relativity, this gravitational curvature is due to the changes in the flow of time only. For fast objects, space curvature also becomes important, and effective weight is approximately given by $W = mg(1+\beta^2)$ [1]. For light, $\beta \equiv v/c = 1$, and hence attraction to light is twice what it would be for slow particles. Thus, the effective radius of curvature for light is $R = c^2/2g$ [2]. Now suppose we again have a spherically symmetric uniform diffuse mass distribution of constant density, ρ. By Gauss' law, gravity at any point only depends on the mass enclosed within a given radius, R. Find R so that the radius of curvature just matches the radius of the body:

$$g = \frac{MG}{R^2} = \frac{4\pi R^3 G}{3R^2} = \frac{c^2}{2R} = g, \quad \textit{so again}, \quad R^2 = \frac{3c^2}{8\pi\rho G} \tag{10}$$

Also, $MG/R^2 = c^2/2R$ very crudely means that $(MG/c^2R) \approx 1$. This is called the "standard cosmological approximation.' Unfortunately, in a "real closed" universe, the concept of a total mass-energy is not well defined– there can be no "outside" observer.Again, this is oversimplified and doesn't use Einstein's field equations for a cosmology. (Nevertheless, it is an interesting coincidence that all these approaches yield the same equation for the radius of the universe).

3. Friedmann Universe:

The derivation of the Friedmann-Robertson-Walker cosmology comes from the Einstein Field Equations. The zeroth terms of these equations is called the "initial value equation" or "I.V.E." and has the form:

$$G_{oo} = \frac{3\dot{R}^2}{c^2R^2} - \frac{3k}{R^2} - \Lambda = \frac{8\pi G}{c^2}[T_{oo} = \rho] \tag{11}$$

The matter on the right tells space-time how to curve on the left side. When divided by 3, the right hand side becomes what is sometimes called the "maximum radius of the universe" $A_o = 8\pi\rho G/3c^2$. In addition, the leftmost term can be expressed in terms of the Hubble expansion, $H = \dot{R}/R$. Then, if the curvature case is $k = 0$ and also if the cosmological constant $\Lambda = 0$, then $H^2 = 8\pi\rho G/3$ so that a critical universal density can be written as $\rho_c = 3H^2/8\pi G$.
There is a special case of a static Einstein cosmology where $\dot{R} = 0$ or $H = 0$. If it were

also true that $\Lambda = 0$ and if curvature $k = +1$, then the equation becomes $1/R^2 = A_o$, or $R^2 = 3c^2/8\pi\rho G$ [3]. So that very special case again matches the above universal radiii.

Expanding closed universes also make use of A_o. One can introduce a radius like the Schwarzschild radius:

$$D = 2m = \frac{2MG}{c^2} = \frac{2(4\pi GR^3\rho)}{3c^2} = A_o R^3 \tag{12}$$

In proper general relativity, the Friedmann metric for a closed spherical static universe gives a universe radius of $R = 4MG/3\pi c^2$. The volume of this universe is the surface area of a 3-sphere embedded in a four dimensional space time, $S^3 \subset R^4$, $A_3 = 2\pi^2 R^3$. So $M = \rho A_3 = 2\pi^2 R^3 \rho$, then:

$$R = \frac{8\pi^2 R^3 \rho G}{3\pi c^2}, \quad or \quad R^2 = \frac{3c^2}{8\pi\rho G}. \tag{13}$$

If we consider $\Lambda \simeq 0$ and $k = +1$, an expanding universe is closed and follows a cycloid growth to its maximum value, the 'Schwarzschild value' D and then contracts [3]. A cycloid is the altitude of a spot on a rotating tire from one contact with a road to the next. The other cosmological cases tend to have unbounded universal radii. For k = 0, the exploding universe continues to expand forever and grows as the 2/3'rds power of time. For the open hyperbolic universe with k = -1, the universe grows to infinity. After Einstein's addition of non-zero cosmological constant for his initially static universe, deSitter in 1917 considered a positive $\Lambda \neq 0$ which yielded a massless universe growing exponentially with time. When $\Lambda > 0$ a $k = +1$ massive universe can still be closed and re-contract as long as the cosmological constant is below a critical value: $\Lambda < \Lambda_c = \sqrt{c^2/4\pi\rho GR^3} = \sqrt{3D/2}$. The future of our own universe is not yet clear. Rather than having a previously anticipated decceleration parameter, expansion actually appears to be accelerating. The cosmological constant may or may not be constant, and its increase or even constancy might drive the universe to infinity.

4. Observable Universe Radius:

The currently most accepted model for our universe is called ΛCDM for a standard expanding concordance big bang cosmology with significant cosmological constant Λ and 'Cold Dark Matter' accompanying the usual baryonic matter and radiation [6]. The big surprise of the new concordance was that the atomic matter density of the universe is well below critical density and that dark matter and dark energy dominate. The density of our universe is $9.3 \times 10^{-27} kg/m^3 \simeq$ an atom of hydrogen per four cubic meters of volume. The total number of atoms in the observable universe is about 10^{80}. This value for density is within a percent or two of representing an expanding flat universe, $k = 0$ and critical density $\Omega = 1.00$, so that $\rho_c = 3H^2/8\pi G$. The composition is now 4.6% atoms, 23% CDM, and $\Omega_\Lambda = 73\%$ dark energy which is now believed to be Λ with negative vacuum pressure. The Hubble constant is $H_{o\,(2010)} = 70.4\; km/s/Mpc = 2.28 \times 10^{-18}/s$. The age of the universe is 13.7 billion years which is also $1/H_o$. [1]

The 'observable radius' is the distance that is observable in principle in the present day in terms of detection of radiation. The earliest light radiation can only be seen back to the time of recombination when photons became free. This is also called the surface of last scattering at an age of about 380,000 years after the big bang and with a ratio $z \simeq 1091$. Those old photons now constitute cosmic black body radiation. It wrong to assume that the actual radius of the observable universe is its age times the speed of light (13.7 ℓy). It is better to say that the 'comoving distance' to those photons is the detectable universe radius of about 46 billion light years [7]. This "proper distance" takes into account the expansion of space during the light travel time. So far (2010), the greatest observed redshift value of a galaxy is about z = 8.55 as seen in the Hubble Ultra Deep Field photos. This looks back in time about 13 Gya or 30 $G\ell y$ distance. The calculation of proper distance can be done by the integral [9]:

$$X = \int_o^z \frac{dz'}{H(z')} \simeq \int_o^z \frac{dz'}{H_o\sqrt{\Omega_\Lambda + \Omega_m(1+z')^3}} \tag{14}$$

A numerical Integration check calculation agrees with the 46 $G\ell y$ proper current radius (for parameters 13.7 billion years, 0.73,0.27 and $z \sim 200$) and also the 30 $G\ell y$ for z = 8.55. The denominator (squared) of this equation (14) is a special way of writing Einstein's I.V.E. equation (11) where the fractions of critical density are $\Omega_{matter} = 8\pi\rho G/3H_o^2$ and $\Omega_\Lambda = \Lambda/3H_o^2$ [10]. Also consider that the baryon density times Radius cubed is a constant and that the red-shift parameter is defined as $(1+z) = R(t_{received})/R(t_{emitted})$.

5. Actual Universe Size:

The actual size of the universe is probably much larger than that which can be observed now or even many billions of years from now. A "horizon" is a universal distance beyond which we cannot see. There are likely many regions with their own "particle horizons" beyond our observational limits at current time, and they could each be defined by a redshift factor $z = \infty$. Many of these regions will never be observable to us even in the far future.

Cosmologists are increasingly accepting the concept of initial big bang "inflation" in which the observable universe starts as a truly tiny causally connected object that balloons out to a large size within 10^{-36} s after the big bang origin. For early inflation with huge Λ and also for later universe growth with large R, the initial value equation (11) has dominant Λ. The metric is also Λ dominated, so:

$$g_{oo} = {g_{rr}}^{-1} \simeq (1 - \Lambda r^2) \Rightarrow \dot{R} \simeq cR\sqrt{\Lambda/3} \Rightarrow R = R_o e^{\sqrt{\Lambda/3}t} = R_o e^{Ht} \tag{15}$$

[1]Recent analysis of the Planck satellite data from March, 2013 [16] states slight changes in cosmic parameters such as $\Omega_\Lambda = 0.693$, $H_o = 68\ km/s/Mpc$, $Age = 13.8\ Gyr$. Baryon density is 4.6% and dark matter increased to 26.8%.

with equation of state $p = -\rho$. The de Sitter Universe expands exponentially to infinity but is also closed with topology $R \times S^3$.

With inflation, the total universe could be 10^{23} larger than the observable universe! The idea of initial inflation solves three previously outstanding cosmology problems: the apparent lack of magnetic monopoles, spatial flatness, and the "horizon problem" – that there should be many causally disconnected regions in the sky. GRT is based on a cosmological principle that requires the universe to be homogeneous and isotropic. Inflation smooths out inhomogeneities and anisotropies and flattens out curvature towards flatness. But most importantly, it is consistent with the subtle temperature variations and their angular scales observed for the cosmic black body radiation (e.g., as seen by the WMAP spacecraft).

6. Empty Universe:

After Einstein published his general theory of relativity, he was surprised that an exact solution could be found so quickly by others and also by its subsequent development by later authors. Even the Schwarzschild solution was a surprise because it gave a solution to a point mass source without regard to the rest of the universe. The distant background is Minkowski space-time. De Sitter's solution had a universe completely without matter just using the cosmological constant. Einstein was gradually becoming suspicious of Mach's principle– one of his initially guiding concepts. Mach's principle says that inertia should be due to the contributions of all the matter in the universe. There should be no absolute frames of reference– only frames of reference relative to distant matter. Later, Kurt Gödel presented his rotating universe solution (1949) whose ties to Mach's principle are ambiguous at best. Even an empty universe $\rho = 0$ has a solution. In general, for a homogeneous isotropic 3-geometry of constant curvature, the Einstein I.V.E. equation (11) implies that:

$$ds^2 = c^2 dt^2 - a^2(t)[d\chi^2 + \Sigma^2 d\Omega^2], \quad with \quad \Sigma = \sin\chi,\ \chi\, or\, \sinh\chi \tag{16}$$

for curvatures $k = +1, 0,\ or\ -1$ and hyperpolar angle χ [1]. A universe with no matter or cosmological constant has $T_{\mu\nu} = 0$ and $\Lambda = 0$ which certainly has density below critical density and should be an open hyperbolic universe. Also, $\dot{R} = cnst$ which could be chosen as the speed of light. The appropriate Robertson-Walker metric can then have universe radius a(t) = ct [12]. This 3-D spatial universe can be pictured using a 4-D embedding diagram with vertical axis $w = (ct') = (ct)\cosh\chi$ and a radius $r' = a(t)\sinh\chi = (ct)\sinh\chi$. The metric is then a hyperboloid of revolution. But the embedding equations can also be considered as coordinate transformations resulting in a flat Minkowski metric for a static spacetime:

$$ds^2 = c^2 dt^2 - [dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)]. \tag{17}$$

By the principle of covariance, either form is equally useful to describe the physics. We could say that the empty universe is a rigid Minkowski frame, K', or an expanding frame K, with an $a(t) = ct$ and with χ, θ, ϕ held constant. Then $ct\chi$ increases as Hubble's law in K but no expansion in K'. So, it could be said that the empty universe does not have a well defined radius. Also note that in our flat real universe, there is ambiguity in the choice of expanding radius versus hyperpolar angle. If we let "s" be a scalar multiplicative factor like 1 or 2 or 3 times a(t), then $a(t)\chi = (as)(\chi/s)$ so that $ad\chi$ and $a\chi$ are preserved for any choice of s. But this is only true for flat $k = 0$ universes and not for $k = -1$ *or* $k = +1$ universes. So, the flat universe also doesn't have a very well defined radius.

7. Problems:

General Relativity has been very successful and is now well accepted for weak field solar system physics and strong field double pulsars, and its limiting case of Newtonian gravitation and mechanics had centuries of success. But since the work of Zwicky (1937) and Rubin (1970), there have been problems using Newtonian mechanics for galactic cluster dynamics and for the rotation rate radial profiles of spiral galaxies. It seems to be necessary to invoke the general existence large heavy dark matter galactic halos to produce the observed flat rotation curves consistently seen in galaxies. Although dark matter also seems to be required for the cosmos too, there is a problem in that dark matter has never been detected in the laboratory. Also CDM often requires fine tuning to fit data.

One possible solution to this "apparent problem" was the transition to a "Modified Newtonian Dynamics" or "MOND" as an alternative to the assumption of dark matter (Mordehai Milgrom 1983 [4]). This is so far only a phenomenological platform currently lacking a theoretical basis and says that the gravitational force law will be modified when accelerations , a, are very low (e.g., a value $a_o \simeq 1.2 \times 10^{-10} m/s^2$). A modification function may resemble $\mu(a/a_o) \simeq (a/a_o)/\sqrt{1 + (a/a_o)^2} \to (a/a_o)$ for weak accelerations. At large distances r, galactic accelerations $a \ll a_o$. Then $GMm/r^2 = F \to ma\mu(a/a_o)$ with $a \simeq v^2/r$ can give flat rotation curves for $v(r)$. There is still growing evidence for the apparent validity of MOND at this level of size. For clusters of galaxies, MOND currently doesn't work quite as well, and GRT requires substantial dark matter. MOND also doesn't fit well with large scale structure and the CMB data. And MOND also isn't relativistic and cannot account for gravitational lensing. But, there are modern relativistic versions such as Bekenstein's "TeVeS" tensor-vector-scalar-theory that yield both MOND and also proper gravitational lensing without using CDM [5]. Newer versions can have a scalar or vector field play the role of dark energy.

John Moffat has also had significant recent success in accounting for galaxy rotation curves and mass profiles of galaxy clusters with his Modified Gravity [8] (MOG, also called STVG for scalar-tensor-vector gravity from his previous non-symmetric gravity theory which in turn was based on Einstein's unifield field theory). No dark matter is required at all. He claims that at the time of release of CMB photons, gravity was stronger than it is now so that $7 \times G_{now} \times 4\%\, baryons \simeq 30\%$ total matter density claimed then [14]. His new theory also accounts for acoustic peaks in CMB and possibly for an accelerated universe expansion. He believes his MOG can also represent the Mach-Sciama origin of inertia as due to induction from very distant matter.
There is also a new "conformal gravity" with an effective potential $\phi(r) = 1 - 2m/r + ar + br^2$ which has two new terms a and b (P. Mannheim, 2005-7). It is possible to give these terms

small constant values which could produce a dark matter galactic acceleration constant and a dark energy.

Physicists are enormously creative. For every new test verifying GRT, new theories are created which become new possible successors. There are reasons to want to have new physics. GRT and quantum mechanics still don't dovetail (and string theory may remain intangibly unprovable). There is no understanding for the existence and magnitude of the cosmological constant. The origin of inertia and Mach's principle are still being discussed and why space-time has the background geometry that it seems to have [11]. Gravitational waves are not yet detected by direct experiment (although a Nobel prize has been granted for indirect proof). There is much reason to desire physics beyond the standard model of particle physics– but only a few glimmers that it may exist (like neutrino mass). A theory of everything would still be desirable.

References

[1] C.W. Misner, K.S. Thorne, and J.A. Wheeler, Gravitation, Freeman, San Francisco, 1973
[2] David Peterson, "Elementary Derivations of the Radius of the Universe," 9/4/79, Unpublished.
[3] Ronald Adler, Maurice Bazin, Menahem Schiffer, Introduction to General Relativity, McGraw-Hill, N.Y., 1965,
[4] Milgrom, M., 1983, Astroph. Jour. 270, three articles at pgs. 365, 371, 384.
[5] Jacob D. Bekenstein, "The modified Newtonian dynamics–MOND and its implications for new physics," ArXiv: astro-ph/0701848v2 27 Mar 2007.
[6] J.P. Ostriker, Paul J. Steinhardt, "Cosmic Concordance," arXiv:astro-ph/9505066v1 16 May 1995.
[7] "Observable Universe," $http://en.$wikipedia.org/wiki/Observable
[8] "Modified gravity and the origin of inertia," J. W. Moffat, V.T. Toth, Mon. Not. R. Astron. Soc. 395, L25-L28 (2009) or at arXiv:0710.3415v6 [gr-qc] 10 Apr 2009 which references a previous arXiv: gr-qc/0506021v7 11 Dec 2005 on STVG theory.
[9] Dorothea Samtleben,Suzanne Staggs,Bruce Winstein, "The Cosmic Microwave Background for Pedestrians: A Review for Particle and Nuclear Physicists," arXiv:0803.0834v1 [astro-ph] 6 Mar 2008.
[10] P.J.E. Peebles, Principles of Physical Cosmology, Princeton University Press, N.J., 1993.
[11] R. Michael Jones, Quantum Gravity and Mach's Principle: resolving the inconsistency between General Relativity and quantum theory by considering the interaction of local physics and distant matter, Dec. 13, 2008, 500 pages, unpublished.
[12] Richard J. Cook, M. Shane Burns, "interpretation of the cosmological metric," Am. J. Phys. 77 (1), January 2009, p.59 or see: arXiv:0803.2701v2 [astro-ph] 3 Sep 2008.
[13] Joseph Silk, The Big Bang, 3rd Edition, W.H. Freeman and Co., N.Y., 1980, 2001, pg. 96
[14] John W. Moffat, Reinventing Gravity, A physicist goes beyond Einstein, Smithsonian Books, Collins, 2008.
[15] Mark Whittle, Cosmology: The History and Nature of our Universe, The Great Courses Guidebook, The teaching company, 2008.
[16] Planck Collaboration, "Planck 2013 results. XVI. Cosmological Parameters," arXiv:1303.5076, 20 Mar 2013.
[17] George Smoot, Wrinkles in Time, 2007, pg. 189.
[18] http://en.wikipedia.org/wiki/Physical_cosmology#Energy_of_the_cosmos.

8. Appendix: Energy of the Universe

Conservation of energy in Big Bang? Is there some calculation showing that negative gravitational energy balances the deSitter expansion? Something this important should have some clarity, but details seem hard to find.

One of the biggest ideas in Cosmology is that the Universe could have been created from nothing and that the negative energy of gravity can balance the positive energy of the universal content so that zero energy is preserved during expansion. Most cosmological models are designed to conserve energy, but the statement above is not found in formal cosmology textbooks! The idea is mainly given in popular sources.

For example, George Smoot [17] writes, "In inflation, the total energy of space minus the gravitational attraction of the other parts of space is essentially zero. Therefore, we can still conserve energy and make everything in the universe starting from practically nothing." The universe can grow to $10^{52}\,kg$ in positive mass and still have zero net energy. Or Guth: "The resolution to the energy paradox lies in the subtle behavior of gravity. Although it has not been widely appreciated, Newtonian physics unambiguously implies that the energy of a gravitational field is always negative, a fact which holds also in general relativity." [Zero energy Free Lunch Universe]. Linde also says that positive and negative energy cancel, and energy is conserved. The discussion of inflation begins with the first law of thermodynamics on pressure versus density.

Formal Texts define energy conservation in terms of the (covariant) derivative of the energy-momentum tensor being zero (but without much further elaboration or discussion of the negative energy of gravitation). In electricity and magnetism, energy density goes as E squared and as B squared. Similarly, in Newtonian gravity, it goes as (g squared), with a minus sign. It isnt perfectly clear where the gravity field is in GRT.

The most intelligent overview is probably this [18]:

There is no unambiguous way to define the total energy of the universe in the current best theory of gravity, general relativity. As a result it remains controversial whether one can meaningfully say that total energy is conserved in an expanding universe. For instance, each photon that travels through intergalactic space loses energy due to the redshift effect (but this is a metric or coordinate effect). This energy is not obviously transferred to any other system, so seems to be permanently lost. Nevertheless some cosmologists insist that energy is conserved in some sense. I interpret this to mean that while in some cases it is not meaningful to say that energy is conserved, in these same cases it is no more meaningful to say that it isn't. The issue is the well-definedness of energy, not whether or not it is conserved.

COSMOLOGICAL DISTANCES

DAVID L. PETERSON

Abstract. Hubble's Law as a linear relationship between universal recession velocity and galactic distance only applies for nearby values of redshift. General discussions of cosmological distances in available literature often seem to be pedagogically weak, surprisingly confusing, poorly defined and even inconsistent. Appropriate and up-dated key concepts are summarized here for convenient reference. The relevant pathway of light through space-time is a fascinating surprise to new students of cosmology. Although light from distant galaxies arrives to us at speed c in our present frame, it often begins by being tossed outwards away from us at high speed the other way. Our knowledge of galaxies is constrained by their worldline intersection with our light cone. That means that our measures are not really of galaxies themselves but of their light rays.

1. Introduction:

A basic introductory concept in cosmology is that of an expanding universe approximated by Hubble's Law, $V = H_o D$: the velocity of galactic recession is proportional to the distance from the observer. The word "velocity" is often a misnomer. The convention for "redshift velocity" is to state a recessional speed that would have caused the same z-redshift value $V_{rs} \equiv cz$ *IF* it were due to a linear Doppler effect. Many sources state that recession is observed via wavelength redshifting but then, like Edwin Hubble himself, mistakenly call this a Doppler effect. It is now established that redshift is mainly cosmological and not a classical or special relativistic effect. Plots of V versus D are often displayed for "Hubble's Law," but both axes are often ill-defined.

There are many types of distances used in cosmology but few sources clearly say what type of distance should be used in Hubble's law. There is subsequent inconsistency about whether superluminal expansion is real and whether expansion creates increasing space between galaxies. Many other basic confusions have been recently discussed by Davis and Lineweaver [1], Ned Wright [2], and by Cook and Burns [3]. The Hubble 'constant' is defined to be $H = H(t) = \dot{a}/a = \dot{R}/R = V/R$, but it is not constant and varies with time. In this equation, R is a universal radius and a is a scale factor, $a(t) = R(t)/R_o$ which is a function of universal time to present time, $f(t/t_o)$. A major goal of the 20th century was to pin down the Hubble constant, H_o defined for the present time, t_o. The idea of $V = H_o D$ only really applies to nearby distances. It is more appropriate to avoid the unmeasurable

---

Date: April 21, 2011.

email: davepeterson137@gmail.com. Paper updated to October 5, 2011, March 24, 2013.

concept of distant speed and redefine the Hubble Law as observed redshift, z, being proportional to the distance from the source (Bruhat). But even this only accurately applies to low values of z.

In the sections that follow, we consider first the presently viewed universe called LCDM for Lambda dark energy plus cold dark matter with parameters measured from analysis of the cosmic black body radiation (CMB), supernovae studies, large galaxy surveys, and mathematical modeling fits. This is followed by a discussion of the previously favored ideal model for a universe containing only matter, the Einstein de Sitter Universe (EdS). This has the benefit of tangible simplicity, closed form solutions, relevance to the early real universe and cosmological flatness. It helps simplify concepts that are otherwise difficult.

The two key cosmological distances being discussed are sometimes called "comoving" distance (D_c or D_o –where the subscript zero means at the present time or "now") and "proper" distance (D_p). A similar distance is $D_A = D_o/(1+z)$ called angular size distance or angular diameter distance of "diameter" distance. $D_p \simeq D_A \simeq D_e$ – the distance at the time of emission. $D_e = D(t_e) = D_o/(1+z) = a(t)D_o$ where a is "scale factor." One other distance often used is "luminosity" distance, $D_L = (1+z)D_o = (1+z)^2 D_A$. In modern cosmology, far distance is often measured in terms of the "distance modulus" $m-M$, where m is the apparent magnitude of the source, and M its absolute magnitude. This distance modulus is related to the luminosity distance by: "DM" $= m-M = 5\log_{10}[D_L(Mpc)]+25$. An older convention was to measure distance in units of 10 parsecs so that with far distances the $25 = 5\log_{10}[100,000]$. Sometimes a correction term, K , is used to express what spectral band is being measured. Then Hubble plots are shown as Log-Log for Log(DM) versus Log(z) which often appears as a nearly straight line.

We would prefer to utilize a distance where Hubble velocity, V, is both HD and also the slope dD/dt. But cosmological distance is defined as a product of a time dependent scale factor, $a(t)$, and a time independent comoving coordinate distance $r = R_o\chi$. So time derivatives can only operate on the scale factor. There are a great many different notations and conventions in the literature. We see back in time using cosmic redshift with parameter z. Galaxies that are seen emitted their light long ago at a time of emission, t_e, after the "Big Bang" (BB). The relevant recession speed is the velocity that existed at that time; but we also often refer to the deduced velocity corresponding to time now, t_o. Cosmological expansion is often presented as dots moving apart on the surface of an inflating balloon or an expanding cloud of dust or an expanding "3-sphere." But radial expansion can be considered simply using a one-dimensional stretching rubber band ruler with uniformly spaced marks on it [4]. The marks or dot-galaxies represent comoving distances, D_c, and can be related to an adjoining fixed ruler's marks where both rulers are attached at their bottom zero-point. The fraction of distances on the rubber band versus those of the rigid ruler is the scale factor, and "proper distance" is the comoving distance times this scale factor. Proper distance is $D_p = aD_e = f(t/t_o)D_c$. At time equals present time, the two measures are the same, $D_p(t_o) = D_c(t_0)$. In the comoving stretching rubber band system,

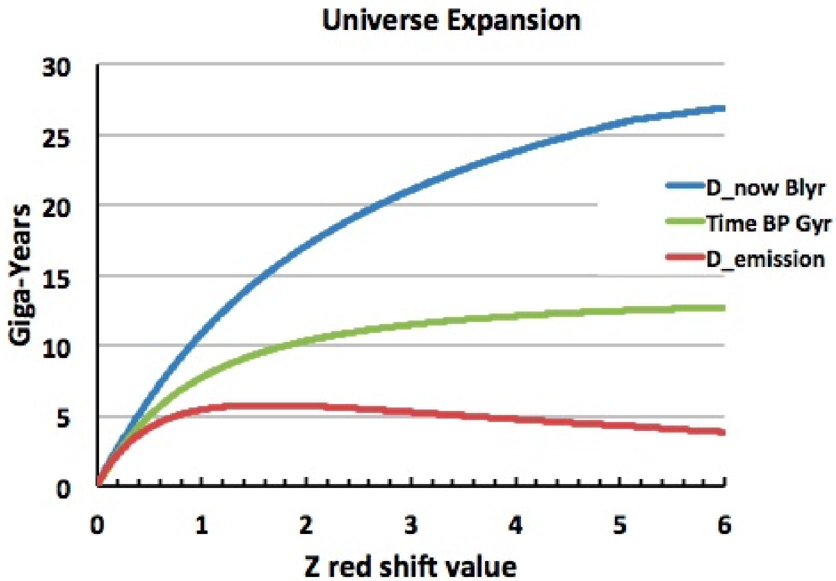


FIGURE 1. Key Distances for Our LCDM Universe

distances between galaxies do not change because the length markers expand along with the cosmic fluid. The fixed ruler measures proper distances.

2. Distances for the ΛCDM Universe:

Distance to galaxies can be estimated by many methods including apparent sizes, luminosity, and type 1a supernovae as standard candles. The most frequent type of distance stated in popular articles is "light travel time" ("ltt"): $D_{ltt} = c(t_o - t_e)$ in light years from time of emission to present time (subscript zero). This is also one of the worst and most useless choices because recession versus this distance is rarely a straight line (Hubble's Law doesn't work for this distance). The appropriate distance is "proper distance" defined as a sum over many intermediate radial distance portions each measured by observers co-moving with the expanding space. But it is somewhat difficult to use and is model dependent. The best defined real measure for distance is simply the cosmological redshift factor "z" or "$1 + z$" for elongation of spectral wavelengths over distance:

$$\frac{\lambda_o}{\lambda_e} = 1 + z = \frac{\nu_e}{\nu_o} = \frac{R_o}{R_e} = \frac{a_o}{a_e} = \frac{1}{a_e} = \frac{1}{S(t)} = \frac{T}{2.7^o K} \tag{1}$$

where ν is frequency, a or S is a "Scale factor" at the time of emission('e) and $a_o = a_{now} = 1$ as a normalization convention. In previous history, the scale factor was a fraction of one. Note that these equations mean that the radius then was $R_e = a_e R_o = R_o/(1 + z)$. T is the temperature of the universe – currently 2.7K.

The calculation of proper distance depends on the cosmological model assumed and being used. These usually come from the Friedmann-Robertson-Walker (FRW) cosmologies which in turn derive from the Einstein Field Equations: $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}/c^2$ where $G_{\mu\nu} = R_{\mu\nu} - g_{\mu\nu}R/2$ ["Ricci tensor and Ricci scalar, R"]. The zeroth terms of these equations is called the "initial value equation" or "I.V.E." and has the form:

$$G_{oo} = \frac{3\dot{R}^2}{c^2 R^2} - \frac{3k}{R^2} - \Lambda = \frac{8\pi G}{c^2}[T_{oo} = \rho] \tag{2}$$

The "k" is curvature (+1 for spherical universes, 0 for Euclidean-flat, or -1 for hyperbolic). The matter term on the right tells space-time how to curve on the left side. It is a frequent convention in general relativity to set $c \equiv 1$ and $G \equiv 1$

Universal density, ρ, is composed of matter (baryons and dark matter), radiation, curvature, and the new dark energy – which could be the cosmological constant, Λ from the old and previously discarded Einstein Universe. The fractions of these with respect to that needed to "close the universe" at present are labeled Omega: $\Omega_m = 8\pi\rho G/3H_o^2, \Omega_r, \Omega_k$, and $\Omega_\Lambda = \Lambda/3H_o^2$. As the universe expands, matter density dilutes volumewise as $\rho_m \propto a^{-3}$. Radiation also dilutes but in addition loses strength by redshifting so that $\rho_r \propto a^{-4}$. Λ is supposed to be constant and doesn't dilute with the expansion of space. If $a = 1/(1+z)$ where $a_o = 1$, then $\dot{a} = -\dot{z}/(1+z)^2$, so the Einstein I.V.E. can be re-written as:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{\dot{R}}{R}\right)^2 = \left(\frac{\dot{z}}{1+z}\right)^2 = \frac{8\pi\rho G}{3} + \frac{k}{a^2 R_o^2} + \frac{\Lambda}{3} \tag{3}$$

Again, since $a_o = 1$, the matter term is:

$$\frac{8\pi G\rho_{mo}}{3a^3} = {H_o}^2\Omega_m(1+z)^3, \quad and \quad \frac{8\pi G\rho_{ro}}{3a^4} = {H_o}^2\Omega_r(1+z)^4 \tag{4}$$

so that,

$$H = \left(\frac{\dot{a}}{a}\right) = H_o E(z) = H_o\sqrt{\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda} \tag{5}$$

E(z) is an expansion factor also known as the "dimensionless Hubble parameter," $E(z) = H(z)/H_o$. In the current universe of interest (Λ-CDM or just "LCDM"), curvature is flat so that $k = 0$ and radiation is negligible well after an age of 60,000 years [9]. The composition is now 4.6% atoms, 23% CDM, and $\Omega_\Lambda = 73\%$ dark energy which is now often assumed to be Λ with negative vacuum pressure. [1] The Hubble constant is $H_{o\,(2010)} = 70.4\,km/s/Mpc = 2.28 \times 10^{-18}/s$. The age of the universe is 13.7 billion years which is also $t_H = 1/H_o$. Notice that setting the matter term equal to the radiation term in equation 5 yields the Scale factor, S, at the important reference time of 'matter-radiation equality.' Before this time, expansion tends to scale as $S \propto t^{1/2}$, and afterwards as $S \propto t^{2/3}$. Matter is the sum

[1] Recent analysis of the Planck satellite data from March, 2013 [17] states slight changes in cosmic parameters such as $\Omega_\Lambda = 0.693$, $H_o = 68\,km/s/Mpc$, $Age = 13.8\,Gyr$. Baryon density is 4.6% and dark matter increased to 26.8%.

of atomic matter and dark matter, while radiation is the sum of photons and neutrinos. $\Omega_m = \Omega_{at} + \Omega_{dm} = 0.044 + 0.23 \sim 0.27$, and $\Omega_r = \Omega_{ph} + \Omega_{nu} = 5.0 \times 10^{-5} + 3.4 \times 10^{-5} = 8.4 \times 10^{-5}$ [9]. So,

$$S = \frac{S^4}{S^3} = \frac{(1+z)^3}{(1+z)^4} = \frac{\Omega_{rad}}{\Omega_m} = \frac{8.4 \times 10^{-5}}{0.27} = 3 \times 10^{-4} \tag{6}$$

The corresponding temperature is $T = 2.7/S \simeq 9000K$, and the time since the big bang is about 57,000 years.

One can now calculate the age of the universe, $t_o = t_{now}$ by integrating back over the change in scale, a. That is, $dt = da/\dot{a}$, so $t = \int(da/\dot{a})$. If the red-shift-factor $rsf = y = 1 + z = a_o/a = 1/a$, then $da = -dz/(1+z)^2 = -dy/y^2$ and $dy = dz$. For limits of integration, $a = a_o = 1 \Rightarrow z = o,\ or\ y = 1$ and scale factor $a = 0 \Rightarrow z = \infty$. Then

$$t_o = \int_o^a \frac{da}{\dot{a}} = \int_1^\infty \frac{dy}{yE(y)} \tag{7}$$

where E(y) means the E(z) expansion factor using $(1+z) = y$. For time at arbitrary z value, $t(z) = \int dz/[(1+z)E(z)]$ from z to ∞. A frequently used time duration is the "look back time", $t_{lb}(z) = t_o - t(z)$. There is a slight adjustment for $z > 1100$ (CMB) because of the strength of radiation at those early times. A plot of look-back-time versus z rises quickly with z and then rolls over to the current 13.7 Gyr age of the universe.

Comoving (fixed present perspective) distance is defined using the previous concept of $dt = da/\dot{a}$:

$$D_{now} = \int_{t_e}^{t_o} \frac{cdt}{a(t)} = \int_{a_e}^{a_o=1} \frac{cda}{a\,\dot{a}} = \int_{a_e}^{1} \frac{cda}{a^2 H_o E(a)} \tag{8}$$

where E(a) is the previous expansion factor E(z) with the $(1+z) = 1/a$ substitutions. This comoving distance is based on the distance that light could have traveled from emission to us.This form is based on equation (5) with $\dot{a} = aH = aH_oE(a)$. The relevant $\Lambda CDM - CMB$ result in the redshift z form is then given by [11]:

$$D = \int_o^z \frac{dz}{H(z)} = \int_o^z \frac{dz}{H_o E(z)} \simeq \int_o^z \frac{dz}{H_o\sqrt{\Omega_\Lambda + \Omega_m(1+z)^3}} \tag{9}$$

Figure 1 shows key distances as a function of cosmic red shift, z. The top curve is $D = D_{comoving} = D_o = D_{now}$ for the presently perceived LCDM universe. The lookback time curve can also be considered as lookback distance $D_{ltt} = D_{LB}$. The lower curve is proper distance $D_p \sim D_e$ and bows over at $z \sim 1.6$ (and $D \sim 5.6$ Glyr). It also turns out to be constrained by a curve that light would follow with respect to us at the present time– the "light cone." For low values of z or for time near the present time, these three measures of distance are about the same. A numerical Integration check calculation agrees with the 46 $G\ell y$ comoving current radius now (for parameters: age $t_o = 13.7$ billion years,

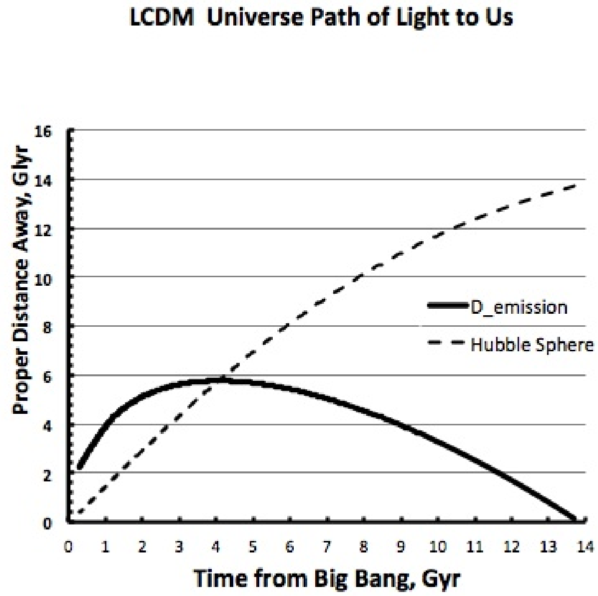


FIGURE 2. Light trajectory from a galaxy at age T to our galaxy at the present time 13.7 Gyr in the "Real Universe." This is the light cone or teardrop shape. The Hubble radius is also shown.

Ω's 0.73,0.27 and $z \sim 200$) and also another heck at 30 $G\ell y$ for z = 8.55 for one of the Hubble Ultra Deep Field photo galaxies. Plots of D(z) and also t(z) agree with available plots found in texts [9] and journals [1]. A plot of D versus lookback time would look very similar in shape and scale to that for the EdS universe shown in Figure 3.

Figure 2 shows the path that light takes in the ΛCDM universe from a distant galaxy to us. Essentially, the light path $\ell_\gamma(t_{forward})$ of photons is at the same proper distance away from us as D_e(time before present) shown in Figure 3 (as $D_e(t_{BP})$). This path is sometimes called a half "tear-drop" shape. There are many nested paths like this, but only this one intersects us at the present time whereas others attain zero distance prior to or later than our time. Also shown is the "Hubble Sphere" distance $c/H = ct_H/E$ dividing regions with recession speeds faster or slower than the speed of light, c. [Relevant points on the curve that agree with a plot from reference [1] are: (z = 1, t = 5.95 Gy, D = 5.44 Glyr), (z = 3, t = 2.3, D = 5.3), and Hubble-Sphere ($z \sim 1.67$, t = 3.97, D = 5.61)].

Some sources prefer to define distance in terms of comoving coordinate or hyperpolar angle χ as proper $D = D_p = R(t)\chi$ or comoving $D = R_o(t)\chi$ where $\chi(z) = \int cdz/[R_o H(z)]$. Then $R_o\chi = \int cdz/H(z)$ like that of equation (9). Analyses based on these angles derive from use of the FRW "cosmological line element"

$$ds^2 = -(cdt)^2 + a(t)[d\chi^2 + \Sigma^2 d\Omega^2] \tag{10}$$

where $\Sigma = \sin\chi,\ \chi\ or\ \sinh\chi$ for curvature $k = +1, 0, -1$ [5]. The geodesic world line for photons that pass between a galaxy to the earth must be radial. The literature is inconsistent on using a as a fraction of unity or as $a(t)R_o$. Note again that this form for metric decomposes radial distance into a time dependent scale factor and a time independent comoving coordinate, χ.

An article from 2007 [14] based on analysis of 182 supernova Ia data show the deduced values for the universal deceleration parameter for the current universe. The strongest statistical case for an accelerating universe is seen at $z \sim 0.2$ and might have a mean expected value near $q = (1 + 3w)/2 \sim (1 + 3(-1))/2 = -1$ at the present time $z = 0$. But for $z > 0.8$, $q \sim +0.5$ as expected for a matter dominated early universe (like that of the Einstein de Sitter model discussed below). The transition from deceleration to acceleration may be fairly recent near $z \sim 0.36$. The dark energy state parameter w may have had a value above -0.6 by $z \sim 2.0$. The current value w = -1 is similar to dark energy being like a cosmological constant. Dark energy is an anti-gravity field with negative pressure but with value varying with cosmic time rather than actually being constant like Λ.

3. Einstein de Sitter Universe:

The "Einstein- de Sitter Universe" (EdS) is an FRW homogeneous and isotropic cosmological model for a "matter-only" flat expanding universe without pressure or cosmological constant and with "just-right" zero spatial curvature (k = 0). This model was proposed by Einstein and de Sitter in 1932 as a "simplest reasonable case" and was highly popular for half a century into the 1980's. Problems began to develop with not being able to find enough luminous matter to close the universe so that it appeared for awhile that the universe might be "open." Despite its historical importance, EdS is not prominent in texts on general relativity or cosmology. Nevertheless, it continues to serve as a highly tangible example lending itself to elementary integrations. In addition, it represents a flat universe – as we have now. And our early universe was matter dominated with weaker dark energy effect and deceleration parameter comparable to that of the EdS model ($q \simeq +0.5$).

The EdS model can be discussed within a Newtonian framework by considering an all matter universe just at the critical rate of expansion for which total energy is $E'_o = 0$. That is, at the present time, let the universal radius = R_o with expansion rate v_o. Energy is always conserved so $v^2/2 - GM/R = v_o^2/2 - GM/R_o = E'_o = 0$. Use $M = 4\pi\rho R^3/3 = 4\pi\rho_o {R_o}^3/3$ and divide by ${v_o}^2$ to get:

$$\left(\frac{v}{v_o}\right)^2 - \frac{8\pi G R^2 \rho}{3{v_o}^2} = 1 - \frac{8\pi G {R_o}^2 \rho_o}{3{v_o}^2}$$

Again, the "scale factor" for the size of the universe is $a = R/R_o$ where present value $a_o = 1$. The critical density for borderline universal expansion at present era is $\rho_{oc} = 3H_o^2/8\pi G$

where present Hubble $H_o = v_o/R_o$. Then KE + PE = Total energy becomes:

$$\left(\frac{v}{v_o}\right)^2 - \frac{a^2\rho}{\rho_{oc}} = 1 - \frac{\rho_o}{\rho_{oc}} = 1 - \Omega_m = 0 \tag{11}$$

The rate of scale $da/dt = dR/R_o dt = v/R_o = (v/v_o)(v_o/R_o) = H_o v/v_o$, then $v(t)/v_o = t_H da(t)/dt$ where extrapolated "Hubble Time" is $t_H = 1/H_o$. So, $(v/v_o) = (\dot{a}/H_o)$. Again, as the universe expands, matter density dillutes volumewise as $\rho_m \propto a^{-3}$ and $a = 1/(1+z)$. Reassembling these into the above equation (11) gives a more familiar form:

$$H = \left(\frac{\dot{a}}{a}\right) = H_o E(z) = H_o\sqrt{\Omega_m(1+z)^3} = H_o\sqrt{\Omega_m/a^3} \tag{12}$$

The real universal time from the Big Bang to now, t_o, can be found in a variety of ways. Since $\dot{a} = da/dt$, $dt = da/\dot{a}$, we again use an equation like (7):

$$t = \int_o^a \frac{da}{\dot{a}} = \int_o^a \frac{da}{aH_o\sqrt{\Omega/a^3}} = \int_o^a \frac{H_o da}{\sqrt{a}} \quad \Rightarrow t = \frac{2t_H a^{3/2}}{3} \Rightarrow t_o = 2t_H/3. \tag{13}$$

where $\Omega_m = 1$ in EdS cosmology. Since this was cleanly integrable, no numerical computations were required for time. The Lookback time before present (BP) is present age of the universe minus a particular time after the big bang (ABB):

$$t_{BP} = t_o - t = t_o(1 - a^{3/2}) = \frac{2}{3H_o}\left(1 - \frac{1}{(1+z)^{3/2}}\right) = \frac{2}{3H_o}(1 - a^{3/2}) \tag{14}$$

An alternative and more Newtonian progression for deriving EdS universal time might be:

$$\frac{da}{dt} = \frac{v}{R_o} = \frac{1}{R_o}\sqrt{\frac{2MG}{R}} = \sqrt{\frac{2MG}{aR_o{}^3}},\ \ a^{1/2}da = \frac{2}{3}da^{3/2} = \sqrt{\frac{2MG}{R_o{}^3}}dt$$

Integrate to get:

$$a^{3/2}(t) = \frac{3}{2R_o}\sqrt{\frac{2MG}{R_o}} = \frac{3v_o t}{2R_o} = \frac{3t}{2t_H}$$

$$a(t) = \frac{R}{R_o} = \left(\frac{3t}{2t_H}\right)^{2/3},\ \ a_o{}^{3/2} = \frac{3t_o}{2t_H},\quad t_o = \frac{2t_H}{3},\ \sqrt{a} = \left(\frac{t}{t_o}\right)^{1/3} \tag{15}$$

Again, this Newtonian result is the same as that obtained by General Relativity. With the current knowledge of H_o from WMAP and other sources, the Hubble time is about $t_H = 1/H_o \simeq 13.4$ Gyr – regardless of the cosmology model. But the age of the EdS universe is only 2/3rds of that or about 9 Gyr. The velocity of the Newtonian universe is:

$$v = R_o\dot{a} = aR_oH_o\sqrt{\Omega_m/a^{3/2}} = \frac{R_o}{t_H}\left(\frac{2t_H}{3t}\right) = \frac{2R_o}{3t_o}\left(\frac{t_o}{t}\right)^{1/3} \tag{16}$$

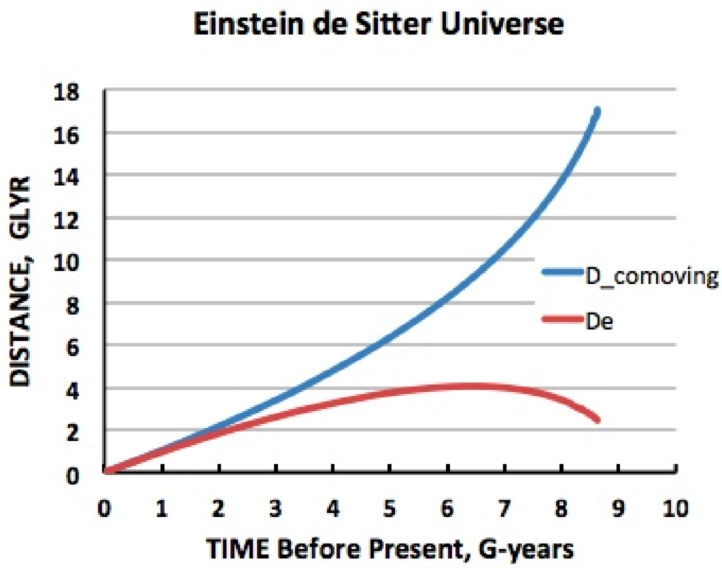


FIGURE 3. Distances versus Lookback Time for EdS Universe. $D_e(t)$, proper distance at emission time, is also the path of light to us.

What is missing from this equation is any reference to some coordinate distance separation between some galaxy and us, $r = R_o\chi$. χ proportions down some maximum distance measure to some relevant distance measure– picking a galaxy or picking a z value. The Newtonian perspective is what an outer shell motion might have for recession at our time t_o. But proper distance equations below are more concerned with the constraint that we have to be observing receding galaxies now. Near $t = t_o$, nearby galaxies must be receding slowly so that velocity is low rather than high. That observation constraint is basic.

The comoving distance, $D_c = D_{now} = D_o$, can be found in terms of scale factor a by using equation (8) or by the usual integration using red-shift z:

$$D_c = \int_o^z \frac{dz}{H_o E(z)} \simeq \int_o^z \frac{dz}{H_o\sqrt{\Omega_m(1+z)^3}} = \frac{2}{H_o}\left(1 - \frac{1}{\sqrt{(1+z)}}\right) = \frac{2}{H_o}(1 - a^{1/2}) \tag{17}$$

Then one can graph D(z) versus lookback time $t_{LB} = t_o - t_e$ from z = 0 to infinity for time from zero to about 9 Gyr, distance values from 0 to about 27 Glyr, and Hubble speeds up to about 3c. [See Figure 3. The curve D(t) for LCDM universe is very similar in shape and scale to this but doesn't truncate until 13.7 Gyr]. The present comoving distance can also be calculated directly from the result of equation (15) for scale a in terms of time, t:

$$D_{now} = \int_{t_e}^{t_o} \frac{cdt}{a(t)} = \int_{t_e}^{t_o} \frac{cdt}{\left(\frac{3t}{2t_H}\right)^{2/3}} = 2^{2/3}3^{1/3}c\,t_H{}^{2/3}(t_o{}^{1/3} - t_e{}^{1/3}) = \tag{18}$$

$$D_o = 3c(t_o)^{2/3}(t_o{}^{1/3} - t_e{}^{1/3}), \quad so$$

$$D_p = D_e = a(t_e)D_o = (t/t_o)^{2/3}D_o = 3c(t_e)^{2/3}(t_o{}^{1/3} - t_e{}^{1/3}) \tag{19}$$

where $t_o = 2t_H/3$. For example, if the red-shift factor for EdS is $rsf = (1+z) = 3$, then $a(t_e) = 1/3$, and $D = 0.845ct_H$ and one can use $t_e = t_H 2/3^{5/2}$. Also $D_{light} = ct = 0.54ct_H$, and $D_{emission} = D_o/(1+z) = 0.28c\,t_H$. [E.g., as in Problem 4 p. 47 of [9]].

If the equation (18) is re-expressed by factoring out the t_o term and using $2t_H = 3t_o$,

$$D_o = \frac{3ct_o}{a_o}\left[1 - \left(\frac{t_e}{t_o}\right)^{1/3}\right] \rightarrow \quad \left[R_{\,horizon} = \frac{3ct_o}{a_o} \sim 27\,Glyr\right] \tag{20}$$

by taking the limit of $t_e \to 0$ back to the origin. This "particle" horizon is the distance that light has traveled from t = 0 to its present time, t_o. If the maximum Hubble speed is stated as a ratio $R_{horizon}/t_o$, then max v = 3c for the EdS. The proper particle horizon is current distance of objects that emitted the oldest light that we can see.

Notice that using weak z values (nearby to us) in equation (17) gives $D \simeq (2c/H_o)(1-(1-z/2)) \simeq 2cz/2H_o = cz/H_o$. When Hubble first measured redshifts, he assumed that they were due to Doppler shifts with $V = cz = H_oD$. Some authors report "cz" versus luminosity distance, D_L and avoid the use of the word "velocity." True Hubble law is really general relativistic expanding space rather than Doppler shift– but the calculations work equally well for nearby galaxies.

In terms of hyperpolar angle, χ, the 'observed' proper distance to a galaxy, G, is given by:

$$\ell_G = a(t_e)\chi_G,\ \frac{d\ell_G}{dt} = V = \dot{a}(t_e)\chi_G = \frac{\dot{a}(t_e)}{a(t_e)}\ell_G = H(t_e)\ell_G. \tag{21}$$

H(t) refers to the Hubble value at the time of light being emitted from a galaxy to us and then being received by us at the present time. So, the Hubble Law pertains to the time of emission of photons, and H is a variable not necessarily H_o. For the EdS model, $H = \dot{a}/a = H_o/\sqrt{a^3}$, and time from BB is $t = 2a^{3/2}/3H_o$, so that $H = 2/3t$:

$$\frac{d\ell_G}{dt} = \frac{2\ell_G}{3t},\ \frac{d\ell_G}{\ell_G} = \frac{2dt}{3t},\ \ell_G \propto t^{2/3} \propto a(t),\ \ell_G(t) = \ell_G(t_e)a(t)/a(t_e). \tag{22}$$

$$\ell_G(t_{now}) = \ell_G(t_e)a(t_o)/a(t_e),\ or\ \ell_G(t_e) = D_e = a(t_e)D_o = D_o/\sqrt{1+z}$$

And by Hubble definition,

$$V = \dot{D} = \dot{a}D_o = H(t)\,aD_o = H(t)D_e$$

In terms of previous discussion, the proper distance measure is D at the time of emission, $D_e = a(t_e)D_o = D_o/(1+z)$.

Einstein's I.V.E. says that $\dot{a} = aH_o(1+z)^{3/2} = aH_o/a^{3/2} = H_o/\sqrt{a}$, which could also be calculated from $a = a(t) = (3H_ot/2)^{2/3}$. Then:

$$V = \dot{D} = \dot{a}\chi = aD_c = \frac{H_o 2(1-\sqrt{a})}{\sqrt{a}H_o} = 2(a^{-1/2}-1) = 2c\left[\left(\frac{t_o}{t}\right)^{1/3} - 1\right] \; or \tag{23}$$

$$\dot{D}(z) = 2(\sqrt{1+z}-1) = HD_e = H_o(1+z)^{3/2}\left[\frac{2(1-1/\sqrt{1+z}}{H_o(1+z)}\right]$$

So, using proper distances, $V(z_e) = \dot{D}(z) = H(z)D_e(z)$.

Hubble's Law with variable Hubble coefficient describes the cosmic flow that carries galaxies along with it. The V used so far refers to recession of galaxies from us as part of the cosmic flow. But we don"t see that directly, we see light. If light is given off by a galaxy, it's speed doesn't go with the cosmic flow and is described by $V_{total} = V_{recession} - c$. The above formula for V differs from formulas below for light by this difference, c.

Consolidating some of the previous integral expressions for distances, recall first that hyperpolar angle chi is defined by:

$$\chi(z) = \frac{c}{R_o}\int_o^z \frac{dz}{H(z)} = \int_o^z \frac{cdz}{R_oH(z)}, \;\; V_{rec}(t,z) = \frac{\dot{R}}{R_o}\int_o^z \frac{cdz}{H(z)} = \frac{\dot{R}(t)R_o\chi}{R_o} = \dot{a}D_o \tag{24}$$

Of course, this can also be transformed and represented using time as a parameter as in equations above (8) and $a = R/R_o$.

$$D_{now} = \int_{t_e}^{t_o} \frac{cdt}{a(t)} = R_o\int_{t_e}^{t_o} \frac{cdt}{R(t)} = R_o\chi. \quad D_e = R\chi = \frac{R\,R_o\chi}{R_o} = a(t)R_o\chi. \tag{25}$$

Recession velocity at the present time would use $\dot{R}(t) = \dot{R}_o$. The literature on cosmological distances is often confusing in using a(t) to mean R(t) instead of scale factor $a(t) = R/R_o$. Galactic $V_{recession}(t,z) = \dot{R}$ assumes that $c\dot{h}i = 0$ – a fixed comoving coordinate accompanying a galaxy when observed today as having redshift z. Current recession is $V = \dot{R}_o\chi$.

It was previously noted that the slope of $dD_o/dt = 1+z$ was excessively and unrealistically high for both the LCDM (CMB) and the EdS universe. The more appropriate calculation of the slope of D_e versus LookBack Time for dD_e/dt has to include the defined division by (1+z):

$$\frac{dD_e}{dt} = \frac{dD_e/dz}{dt/dz} = \frac{d[D_o/(1+z)]}{dt} = d\left[\frac{2(1-1/\sqrt{1+z}}{H_o(1+z)}\right] /d\left[\frac{2(1-1/(1+z)^{3/2}}{3H_o}\right] \tag{26}$$

$$= c(3-2\sqrt{1+z}).$$

This function agrees with the lower curve shown for D_e proper distance in figures 3 and 4. For example, $V(z=0)=c$ approaching us. However, this dD/dt slope does not match the Hubble velocity above (23). This seems to be one of those occasions where $dD/dt = (dD/dz)/(dt/dz)$ doesn't apply because D consists of a time dependent scale factor part and a time independent coordinate – but both are functions of z. A time derivative only applies to the first part. The discrepancy is that $D_e(t_e)$ is a proper distance lying on the light cone of the light path to us– it is galactic distance constrained by this requirement. The real galaxy is of course continuing to move away from us, but the proper distance is that deduced long ago by the light cone path. In a sense, there are two distances– one continuing outwards (but whose light isn't seen by us at the present time) and one snap-shot long ago on the light cone. The above function is a light path function not reflecting the actual motion of the galaxy even at the time of emission.
As an example, consider a point in time where z = 3 and a(t) = 0.25, or $t = t_0/8$, the velocity of the galaxy is $V = 2(1/(0.25)^{1/2} - 1) = 2(2-1)c = 2c$ outwards. But the above function for $dD_e/dt(a_e) = +1c$. This means that the universal flow out at 2c tosses light back at 1c still away from us. The light then traverses the universe to us, bows over, and arrives at speed -1c.

Most expanding universes have a decreasing Hubble constant because of a positive deceleration value. This is true for the EdS model:

$$q \equiv -\frac{\ddot{a}a}{\dot{a}^2} = -\frac{\ddot{R}}{RH^2} = \frac{\Omega}{2} = \frac{1}{2} \tag{27}$$

That is, substituting a = a(t) from equation (15) into this definition gives the constant $q = 1/2$ which slows down the rate of expansion of the EdS matter universe in which $\Omega_m = 1$.

Older texts referred to a "distance-redshift relation" using the deceleration parameter, q [5] [7].

$$z \simeq H_oD + \frac{1}{2}(1+q_o)(H_oD)^2 + ..., \quad or\ D \simeq T_{LB} + \frac{H_o}{2}(T_{LB})^2 + ... \tag{28}$$

But these estimates for D or for z are already too low by nearly 10% by redshift $z \simeq 0.8$. Since supernovae now go out well beyond $z \simeq 1.0$, these old approximations now have low interest. But, a major goal of old cosmology was to try to experimentally determine the deceleration parameter. The new cosmology no longer shows any current decceleration and, to the contrary, is much more concerned with pinning down the value of cosmological

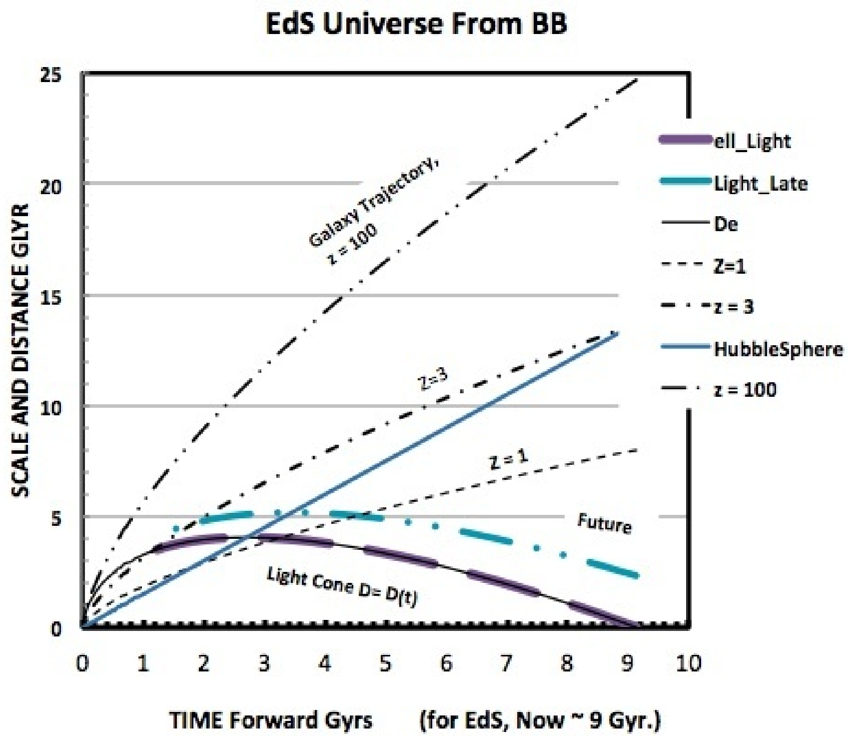


FIGURE 4. Distances versus forward Time from Big Bang for EdS Universe

acceleration.

4. LIGHT RAYS IN EDS UNIVERSE:

Also of interest is the pathway for light travel from distant galaxies. In this case, the speed of light toward us is treated as a "peculiar" velocity so that the net light speed is $v = v_{recession} - c = V_{Galaxy} - c$. Unlike galaxies, light isn't constrained to the 'cosmological fluid.' Consider the Einstein de Sitter Universe in forward time from the Big Bang as shown in Figure 4 for proper distance $D_p = a(t)D_o$ [15]. The time of emission is now some forward time of interest, and present time now in this EdS universe is some 9 billion years from the Big Bang origin with extreme distances out near 27 billion light years $= 3t_o = 3(2/3H_o)$. Forward comoving distance versus forward time is found from equation (18) from time zero to time of emission:

$$D_{BBcomov} = \int_o^{t_e} \frac{cdt}{[a = (t/t_o)^{2/3}]} = c(2^2 3t_H^2 t)^{1/3} \simeq c(12 \cdot 13.7^2\, t)^{1/3} = 3ct_o{}^{2/3}t_e{}^{1/3}. \quad (29)$$

For the present time, this distance is the particle horizon distance, $3ct_o$– the farthest we could possibly see at the present time (as the limit in equation (20)). This equation is just the time complement of comoving distance with respect to lookback time. The scale factor increases as time to the 2/3 power as in equation (15). But again, this distance is missing a reference to some particular galaxy distance, χ.

Galaxies tend to move with the expanding cosmic fluid so that $\chi \sim$ constant while the scale factor and universe expands. The motion of galaxies was discussed previously in formulas (21) and (22). Also the Hubble 'constant' $H = H_o\sqrt{\Omega_m/a^3} = H_o 2/(3H_o t) = 2/3t$. And, for light, Null $ds^2 = 0 = -c^2dt^2 + a^2(t)d\chi^2$ so that $a(t)d\chi/dt = \pm c$. Since light is coming towards us, pick the minus sign for light speed. Then, the key differential equation for light rays from a galaxy to us is given by:

$$\frac{d\ell_\gamma}{dt} = \chi\frac{da}{dt} + a\frac{d\chi}{dt} = H(t)\ell_\gamma - c = \frac{2\ell_\gamma}{3t} - c = V_{tot} = V_{recess} - V_{pec} \tag{30}$$

from some time $t = t_e$ where $\ell_\gamma = \ell_G$ to time now where the photon arrives where we are so that final distance is $\ell_\gamma = 0$ [3]. But the integration of this equation is no longer as trivial as it was for equation (22). The speed of light from the galaxy ultimately towards us may well be initially away from us.

This equation (30) is of the form $\ell' + \ell\, p(t) = g(t)$ so that $p(t) = -2/3t$ and $g = -c$. A solution can be found by using an integration factor $\mu(t) = exp(\int p(t)dt = \exp(\ln(t^{-2/3})) = 1/t^{2/3}$. The initial condition I.C. is $\ell(t_e) = \ell_G$. Then a solution is:

$$\ell_\gamma = \frac{\int(g = -c)dt/t^{2/3} + K}{1/t^{2/3}} = -3ct + Kt^{2/3}. \quad K = \frac{\ell_G}{{t_e}^{2/3}} + 3c{t_e}^{2/3},\ or, \tag{31}$$

$$\ell_\gamma(t) = \ell_G\left(\frac{t}{t_e}\right)^{2/3} - 3c(t - t^{2/3}\,{t_e}^{1/3})$$

An example plot of this result is shown by the lower curve in Figure 4 for a galaxy near the edge of the observable universe when it was very young and now being observed by us in the Milky Way [parameters t age 1.142 Gyr after BB, proper distance $\ell \simeq 3.425 Gly$, scale 0.25, $z = 3$]. The initial speed is away from us at light speed c and the final speed is to us at light speed c. Again, this "light-cone" shape is called "teardrop." If a distant galaxy at the edge casts off its light at a later time, it will not reach us yet but rather further in the future [e.g., age $t = 1.457$ Gyr, distance 4.4 Glyr – also shown in Figure]. At EdS "present time" near 9 billion years after the big bang, the light is still 2.3 billion years away from us and will only intersect us billions of year later. A great many galaxies we see now are much closer than these (well after BB) so that the proper distance line is much shallower than slope 3c They will have light following similar curves to us. The teardrop light cone for the LCDM Universe was shown in Figure 2. The lower curve also shows the overlap between the forward light or photon γ profile and the previous backwards D_e plot versus "look-back" time – both are proper distances. The light cone is the locus of all observable galaxy distances and times labeled by redshift, z. The galaxy itself is traveling an outward path which intersects the light cone at the z-label, but the out-and-up path itself has z varying along its length. Possible examples are shown here. Since Newtonian velocity goes as $(t_o/t)^{1/3}$ power, distance will go as $D \propto \int vdt \propto (t^{+2/3})$ power. So an outward path could be $D_t = D_z(t/t_z)^{2/3}$ where $t_z = t_o/(1+z)^{3/2}$ and $D_t = D_e(t)$ is simply the usual proper distance or light cone curve above. Three of these sample trajectories are shown in Fig. 4 for $z = 1$, 3, and 100. Light from these outward bound galaxies is only seen by us when the galaxy trajectory intersects the light cone.

The final condition of relevant ℓ_γ [eqn. (31)] at the present time is constrained by our observation (no remaining distance) to be:

$$\ell_\gamma(t_o) = 0 = \ell_G \left(\frac{t_o}{t_e}\right)^{2/3} - 3c(t_o - t_o^{2/3}\, {t_e}^{1/3})$$

Solving this condition for ℓ_G constrains the observed location of a galaxy $\ell_G(t_e)$ to be

$$\ell_G(t_e) = \left(\frac{t_e}{t_o}\right)^{2/3} 3ct_o^{2/3}(t_o^{1/3} - t_e^{1/3}) = 3c\,{t_e}^{2/3}({t_o}^{1/3} - {t_e}^{1/3}) = D_e(t_e). \tag{32}$$

So, the space-time path of light from galaxies to us shows distance versus time to be the same for $\ell_G(t_e)$ and $D_e(t_e)$. The only galaxies we see now must lie on our light cone from now back into time. The galaxies themselves continue to advance outwards into space, but we see them when they had a proper distance which was closer and constrained by the light cone.

Alternatively, in terms of just scale factor as a variable,

$$\ell_\gamma(a_o) = 0 = \ell_G/a_e - 3c(t_o - t_o\sqrt{a_e}) \Rightarrow\ \ell_G(a_e) = 3ct_o(a_e - {a_e}^{3/2}) = D_e(a(t_e))$$

The slope of this proper distance ℓ_G is the effective speed of light with respect to forward time, t.

$$d\ell_G(t_e)/dt_e = 2c(t_o/t_e)^{1/3} - 3c = -c(3 - 2/\sqrt{a_e}) = -c(3 - 2\sqrt{1+z}) = -dD_e/dt_{LB} \tag{33}$$

with slope to us of -c but initial slope that is unbounded positive (multiples of c). Note again that the difference between this equation (33) and the cosmic flow of galaxies is $V_{total} = V_{recess} - c$ or $c(2/\sqrt{a} - 2) - 1c = c(2/\sqrt{a} - 3)$.

5. Conclusion:

Exercises: Most of the results here are attempts at understanding by a new student of cosmology, and most of the results above are probably well known by cosmologists. But it seems strange that this knowledge and relevant calculations were so difficult to obtain –it should be clear and consolidated somewhere. That was a purpose of this note. This is also a new exercise in using LaTeX typesetting and attaching figures [grabbing portions of Excel charts as .tiff files, converting to .jpg files and then "includegraphics"]. Excel with many rows was used for integrations since "Octave" (affordable MatLab) and free-GnuPlot won't yet play together on Mac.

Hubble's Law: Writing $V = \dot{a}\chi = H(t)d_p$ makes Hubble's law automatically true– but H is not a constant as originally intended, and neither velocity nor proper distance is measurable. Plotting Hubble's law as just a distance versus redshift, z, curve makes more sense. And new cosmology measurements of luminosity distance versus redshift are also curves but are usually plotted on Log-Log plots which makes them often seem to be linear. Expansion occurs between clusters of galaxies, but atoms and clusters themselves do not expand with time.

Literature results: It was satisfying to obtain some of the same plots provided by cosmology literature (like the distance versus redshift graphs in Whittle [9] and some of the teardrop plots from Davis and Lineweaver [1]). The movement of light at super-luminal speeds away from us seems clear for the early universe. The hardest lesson from this exercise was that we are not mainly considering galactic motions as they really might be but rather subject to the hard constraint that we are observing light from distant galaxies at the present time. The primary observable is redshift, and other properties are deduced from that coupled with presumed models of the universe. The article and dissertation by Tamara Davis is excellent, but the key plots are so small that I had to blow up a one square centimeter section to measure and validate my calculations for Figure 2. No hints on their calculations were ever given anywhere.

Problems: Dark matter has yet to be confirmed by direct measurement – the latest WIMP detection experiments are all failures so far. There is no understanding of what dark energy could be nor why it seems to be changing with cosmic time. It suffers from the so-called fine tuning problem and the cosmic coincidence problem (why is it just right at this time). Although the concordance model is supported by evidence, there are a great many possible alternative models of the universe [16] (Moffat theories [like MOG], Brans-Dicke scalar-tensor, Mannheim conformal gravity, Bekenstein (TeVeS) relativistic MOND, Fourth-order cosmology [f(R)], ...).

References

[1] Tamara M. Davis, Charles H. Lineweaver, "Expanding Confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe," arXiv:astro-ph/0310808v2 13 Nov 2003. More detail is found at Davis PhD thesis: http://www.dark-cosmology.dk/ tamarad/papers/thesis_complete.pdf.

[2] Ned Wright, UCLA, "Fact and Fiction in Cosmology," http://www.astro.ucla.edu/∼ wright/CMB-MN-03/06May05-CSUN-clean.pdf –also see home page for cosmology tutorial and papers, http://www.astro.ucla.edu/ wright/intro.html .

[3] Richard J. Cook, M. Shane Burns, "interpretation of the cosmological metric," Am. J. Phys. 77 (1), January 2009, p.59 or see: arXiv:0803.2701v2 [astro-ph] 3 Sep 2008.
[4] Russell Akridge, "A simple cosmology: General relativity not required," Am. J. Phys. 69 (2), February 2001, p 195-200, or web: http://www.df.uba.ar/users/sgil/physics_paper_doc/ papers_phys/cosmo/simple_cosmology.pdf.
[5] C.W. Misner, K.S. Thorne, and J.A. Wheeler, Gravitation, Freeman, San Francisco, 1973
[6] Ronald Adler, Maurice Bazin, Menahem Schiffer, Introduction to General Relativity, McGraw-Hill, N.Y., 1965,
[7] P.J.E. Peebles, Principles of Physical Cosmology, Princeton University Press, N.J., 1993.
[8] Joseph Silk, The Big Bang, 3rd Edition, W.H. Freeman and Co., N.Y., 1980, 2001, pg. 96
[9] Mark Whittle, Cosmology: The History and Nature of our Universe, The Great Courses Guidebook, The teaching company, 2008.
[10] David L. Peterson, "The Radius of the Universe," LaTeX unpublished, Feb. 27, 2011.
[11] Dorothea Samtleben,Suzanne Staggs,Bruce Winstein, "The Cosmic Microwave Background for Pedestrians: A Review for Particle and Nuclear Physicists," arXiv:0803.0834v1 [astro-ph] 6 Mar 2008.
[12] Tom Theuns, Physical Cosmology, (Antwerp course text on-line). http://icc.dur.ac.uk/∼ tt/ Lectures/UA/L4/cosmology.pdf
[13] Einstein de Sitter Universe for Engineers, http://www.einsteins-theory-of-relativity-4engineers.com/ support-files/einstein-de-sittery.pdf
[14] Yungui Gong, Anzhong Wang, "Reconstruction of the deceleration parameter and the equation of state of dark energy," arXiv:astro-ph/0612196v3 13 Feb 2007.
[15] Excel calcs at: Scratch_1_2.xlsx, Horizon and EdS tabs, dp, 5/5/2011.
[16] Timothy Clifton, Alternative theories of Gravity, dissertation, http://arxiv.org/pdf/gr-qc/0610071v1 13 Oct 2006.
[17] Planck Collaboration, "Planck 2013 results. XVI. Cosmological Parameters," arXiv:1303.5076, 20 Mar 2013.

COMMENTS ON INFLATION

DP

Abstract. The theory of an inflation epoch near the birth of the universe is not only strange and difficult but also incomplete. Discussions of inflation tend to be presented either in hand-waving text or in fairly deep mathematics with not much inbetween. The purpose of this note is to state some of the concepts that make inflation more intuitive, present more elementary mathematics where applicable, and try to resolve frequent stumbling points to understanding. Recently, inflation received strong justification from firm detection of B-mode polariztion seen from telescope at the South Pole [16] .

1. Introduction:

The pre-inflationary standard big bang theory of the universe is unable to explain key problems of cosmology (called the 'flatness' problem, 'horizon,' 'monopoles' (or 'relic abundances'), and 'roughness' (or 'structure')). In 1981, Alan Guth [1] addressed these issues by proposing the existence of a primordial grand unified theory (GUT) phase transition coupled with supercooling and release of latent heat in the earliest history of the universe. He presumed the continued appropriateness of the Robertson-Walker metric and the Einstein field equations of the general theory of relativity (GRT). As a new idea, his early theory had flaws which were then addressed in 1984 by Andrej Linde and Paul Steinhardt enabling the formation of cosmic bubbles of vacuum larger than the presently observable universe [2]. There are now many subsequent possible versions of Inflation to consider[9]. As of 2013, there have been at least 4000 papers written on inflation, and there are also presently at least a few hundred different scenarios for inflation [11] making it a difficult theory to falsify. Categories include models with single-inflation inflation, multiple-field inflation, and non-scalar fields such as vector inflation.

Modern discussions state that near the beginning of universal expansion, there was an epoch of vacuum energy domination, a 'de Sitter' phase or a time when universal pressure was negative. If this inflation epoch lasted long enough, then particle horizons are eliminated, and the sky is homogeneous in all directions from earth. "The most spectacular achievement of inflation is that, combined with quantum mechanics, it provides a convincing mechanism for the origin of the cosmological fluctuations (the seeds of the galaxies and of the Cosmic Microwave Background − CMB − anisotropies) and predicts that their spectrum should be scale invariant (i.e., equal power on all spatial scales) which is fully consistent with the observations [11]. This is particularly interesting because it combines

Date: November 6, 2011.

email: davepeterson137@gmail.com. Paper updated to 8 April, 2014.

quantum mechanics with general relativity.

The concept of spontaneous symmetry breaking (SSB) is basic to modern particle physics theory and is proposed to extend into the very high energy domain of early inflation. The archetypal example of an SSB is the quartic (or 'Mexican Hat') potential $V = V(\phi)$:

$$V(\phi) = -\frac{1}{2}m^2\phi^2 + \frac{1}{4}\lambda\phi^4 + cnst \simeq \frac{\lambda}{4}(\phi^2 - \sigma^2)^2 \tag{1}$$

where σ is the value of the field ϕ giving a minimum value of the potential, V. Its value is $|\phi| = \sigma = m/\sqrt{\lambda} \simeq 246\,GeV$ $-$ a non-zero value $-$ a condensate of Higgs particles. The latter form in equation (1) includes an added constant value on the Mexican Hat so tht V(min) = 0 (for shape, see Figure 1 below). The symbol 'phi' may represent a scalar field such as the most famous originally conceived 'Higgs' field for the breaking of electro-weak symmetry (into massless photons γ, and massive bosons $Z_o, W^+,\ W^-$). A more general phi could represent an effective energy density and pressure of a homogeneous scalar field. Then the potential energy V represents an internal energy corresponding to values of ϕ. For the modern concept of inflation (which occurs well before the electro-weak symmetry breaking scale), the potential has to be modified to give a more prolonged 'slow roll' downhill from the central hilltop. This allows adequate time for inflation to do its job. But the energy scale for primordial inflation is much higher than for electro-weak symmetry breaking (perhaps 10^{15} GeV or so. The shape of the potential is unknown, but popular articles tend to show a (1972,1981) "Coleman-Weinberg" plot inspired by SU(5) GUT model[8]. This shape is shown in Figure 1 and starts out flatter than the rolling Mexican Hat potential. [1] This 'slow-roll single field' model has a slowly moving scalar field is sufficiently flat that the corresponding pressure is negative.

The archetypal "Hat" shaped or "Higgs-type" potential above is not really needed and may now even be disfavored (2014). Instead of falling down a symmetry-breaking potential from a hill at the origin, one can instead start at higher scalar field ϕ (say off to the right) and then fall towards a settling valley at $\phi = 0$. A popular model was Linde's [15] quadratic chaotic inflation or "eternal inflation with Lagrangian having a kinetic term and a parabolic potential density term:

$$\mathcal{L} = \tfrac{1}{2}(\partial\phi)^2 - V(\phi) \quad with \quad V(\phi) = \tfrac{1}{2}m^2\phi^2. \tag{2}$$

This simple view is presently favored by the 2014 results from BICEP2 [16]. The old concepts of supercooling or tunneling from false vacuums is no longer needed for polynomial $V(\phi) \sim \phi^n$ potentials. Linde showed that these potentials are consistent with eternally reproducing universes, and the slope near the origin is sufficiently flat so as to have a slow-roll regime. A simplified heuristic model for multiple universes from eternal inflation is the famous "Cantor" set where the unit interval of all ones is successively modified by deleting middle thirds from false vacuum to universe vacuum. Each new middle third is a bubble universe or pocket universe. The difference is that the original unit length interval

[1] Plotted in Excel "SamplePhysics.xlsx," dp, 11/13/11, Reference [8] Guth, pg. 731.

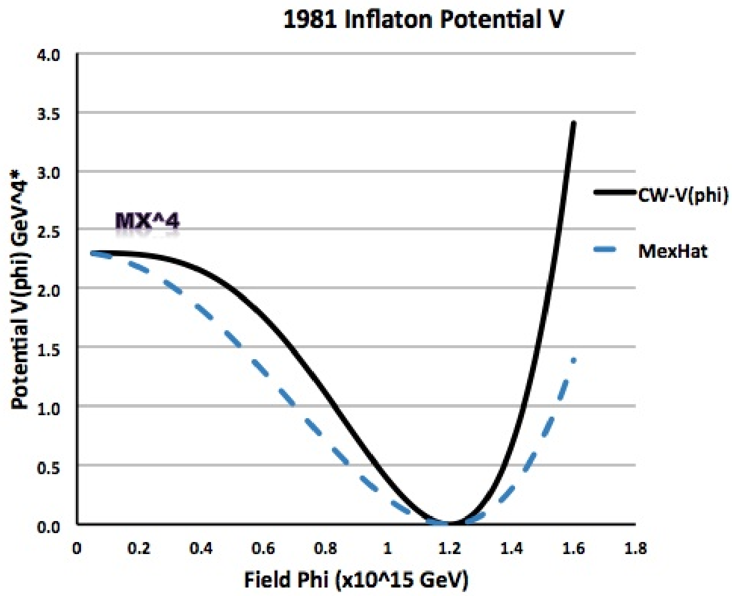


FIGURE 1. A standard shape is the Coleman-Weinberg potential useful for SU(5) symmetry breaking. Also shown is the older electroweak "Mexican Hat" potential for comparison. The vertical scale is $\times 10^{58} GeV^4$ or the mass of an 'X' particle, M_X^4. The minimum is at $\sigma = 1.2 \times 10^{15}$ GeV. $V(\phi) = (25\alpha^2/16)[\phi^4 \ln(\phi^2/\sigma^2) + (\sigma^4 - \phi^4)/2]$, $M_x = 4 \times 10^{14} GeV$, $\alpha = 1/45$

is expanding very rapidly while the deletions take place, and the process is chaotically random rather than regular (as in the traditional Cantor set). The Big Bang commences at the moment when the temperature of the universe reaches its maximum value, right after the end of inflation. The value for the tensor/scalar r ranges from 0.14 to 0.28 for the ϕ^2 *to* ϕ^4 polynomial chaotic inflation models and thus includes the present $r \sim 0.16 - 0.20$ near the quadratic form. The quadratic form also yields n = 0.97 for tilt of power spectra. After further verification of BICEP2 type swirls in other parts of the sky, the next step will be to look for the tilt of the amplitude of the tensor (gravitational) perturbations.

2. Some Basic Math:

Some concepts in general relativity and inflation can also be addressed by elementary Newtonian theory.

Friedman Energy Ballance Equation: Consider a mass being tossed outward from a gravitating body as now being part of a spherical shell of matter being tossed outward and assume that its total energy is conserved so that $KE + PE = E_o$. Let $v = \dot{R} = HR$ where

H is like a "Hubble parameter." Then, per unit mass of shell,

$$\frac{v^2}{2} - \frac{MG}{R} = \frac{H^2R^2}{2} - \frac{4\pi\rho GR^2}{3} = {E_o}' \tag{3}$$

The 'Gaussian curvature' of a spherical surface is given by $K = 1/R^2$ or $KR^2 = +1 = k$. So, with some advanced knowledge, let the total energy term be replaced by $kc^2/2$ [2]. Then,

$$H^2 - \frac{kc^2}{R^2} = \frac{\dot{R}^2}{R^2} - \frac{kc^2}{R^2} = \frac{8\pi G\rho}{3} \tag{4}$$

This equation is very similar to the Einstein general relativity field equation (5). That is, The zeroth term or "initial value equation" or "I.V.E." or the G_{oo} Einstein Equation has the form:

$$G_{oo} = \frac{3\dot{R}^2}{c^2R^2} - \frac{3k}{R^2} - \Lambda = \frac{8\pi G}{c^2}[T_{oo} = \rho], \tag{5}$$

which now includes the cosmological constant term, Λ. The matter on the right tells space-time how to curve on the left side.

Newtonian Fluid Expansion: A differential form of Newton's classical gravitation can be extended into GRT by the inclusion of pressures in the cosmological case where pressure has values competitive with density [5]. This is a simple consequence of energy having mass-equivalence so that it gravitates, and pressure contributes to internal energy or 'stress-energy.'

$$\nabla \cdot g = -4\pi G(\rho) \to \nabla \cdot g = -4\pi G(\rho + 3p) \tag{6}$$

Solving this for a spherical mass considered at a radius, R, the gravitational mass is $M = (4/3)\pi(\rho + 3p)R^3$, and the acceleration due to gravity at that radius is:

$$\ddot{R} = g = -\frac{GM}{R^2} = -\frac{4}{3}\pi G(\rho + 3p)R \tag{7}$$

This is second of the Einstein field equations $-$ a fluid acceleration equation. Also note a standard classical fluid equation is given by internal energy U as pressure times volume V

$$dU = d(\rho V) = -pdV = \rho dV = \rho dV + V d\rho,\ d\rho = -p\frac{dV}{V} - \rho\frac{dV}{V},\ \dot{\rho} = -(\rho + p)\frac{\dot{V}}{V} \tag{8}$$

Now, the volume of a sphere is proportional to R^3, so

$$\frac{\dot{V}}{V} = \frac{3R^2\dot{R}}{R^3} = \frac{3\dot{R}}{R},\ \textit{so}\ \ \dot{\rho} = -3(\rho + p)\frac{\dot{R}}{R} = -3H(\rho + p). \tag{9}$$

This 'continuity equation' is consistent with the Friedman and acceleration equations (4)(7) and could also be deduced from them.

Inflation as Accelerating Expansion: The accelerating expansion due to inflation can also be considered simply. The equation (7) applies 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. At any other radius away from center, the mass of the earth that counts towards attraction is due to the mass inside a spherical Gaussian surface at that radius. Near the center, that volume is tiny so that there is little force. 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 $F = -kR = m\ddot{R} = ma$. The period of oscillation is $\tau = \sqrt{3\pi/\rho G} \simeq 1.4$ hours (where average earth density is 5.52 g/cc and, for simplicity, the hole in the earth is presumed to be in vacuum).

Inflation with a huge cosmological constant and with $p = -\rho$ would end up with a net negative -2p anti-source causing effectively a repulsive gravity which makes the universe 'fall outwards.' We consider a spherical shell of 'balls falling outwards.' This form has a repulsive force $F = +kR$, a similar but different differential equation. Instead of sine-wave motion, the solution this time is instead an exponential expansion, $R(t) = ke^{bt}$ where $b = \sqrt{8\pi G\rho/3}$. This is inflation. Two problems are, "how does it start and how does it end?" At tiny time t, when the universe is just a little 'off-center,' the outwards acceleration is also tiny and velocity is tiny. This is in contrast to the usual big bang massive universe where velocity near time = 0 is almost infinite. This difference is required for homogeneous communication avoiding the horizon problem.

Another approach to this expansion is the following: For early inflation with huge anti-gravity cosmological constant Λ and also for later universe growth with large R, the initial value equation (5) has dominant Λ. So:

$$\dot{R} \simeq cR\sqrt{\Lambda/3} \Rightarrow R = R_o e^{\sqrt{\Lambda/3}t} = R_o e^{Ht} \tag{10}$$

with equation of state $p = -\rho$ (or 'w' $= p/\rho = -1$). This de Sitter Universe expands exponentially to 'really huge size,' but, along the way, it is closed with topology $R \times S^3-$ still spherical curvature but with 'really big' radius so that the final curvature is essentially nil or 'flat' Euclidean space.

The universe expands keeping constant density and produces increasing net mass growing from 'nothing.' This increasing mass is a free lunch due to counterbalancing negative energy gravitation. The sum of the growing mass and gravity of the universe could maintain near zero net energy. A grand unified scale could be $\rho \sim (2 \times 10^{16})^4$ [in units where $G = 1$, $c = 1$] so that the time constant is $\sim 10^{-36}$ seconds over perhaps 60 e-foldings to perhaps 10^{-32} sec. [that is t $\sim 1/\sqrt{(10^{16})^4} \sim 10^{-32}$ sec]. Because the vacuum density is held constant near critical values during inflation, the Hubble parameter is also a constant ($H^2 = 8\pi G\rho_c/c$). Then the Hubble radius, $R_H = c/H$, is also constant. The expanding universe contains myriad little Hubble radius volumes inside it.

Friction during Expansion: An analogy to the 'inflaton' rolling down a hill is that it experiences 'friction' as if it were immersed in a fluid. One example from classical physics is the case of a resisting force proportional to velocity − as in Stoke's law for tiny falling bodies and leading to a 'terminal velocity' for tiny droplets. $F = ma = mg - k'v$, for some constant, k'. Or, since gravitational force can be expressed as a slope in potential, $-m\Delta V/\Delta y$, the vertical height y obeys:

$$\ddot{y} = -\frac{dV}{dy} - k\dot{y},\ or\ \ddot{y} + k\dot{y} + \frac{dV}{dy} = 0 \tag{11}$$

In general relativity for fields, the corresponding equation will be $\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0$ (where 'prime' means derivative with respect to phi − derived in equation (15)). The middle $3H\dot{\phi}$ 'friction' term is due to 'the red shifting of the momentum of the field phi by the expansion of the universe [3].' Some authors refer to the initial development of the field as being in 'molasses' [4]. Andrei Linde says that a high friction term is like the scalar field moving slowly like a ball in a viscous liquid[8]. But this is just equivalent to saying that the initial 'slow roll' has a very gradual slope $\Delta V/\Delta\phi$.

Oscillation Period: So, the field moves gradually away from the initial zero value ('slow roll' motion) and then more rapidly falls into the potential minimum where it now oscillates back and forth. A oscillating field tends to lose energy by creating pairs of elementary particles. In general the period of oscillation in a gravitational field is similar to $\tau \sim 1/\sqrt{G\rho}$. There are several derivations of this concept, the simplest of which is in Mark Whittle's course on Cosmology [4]. As a particular example, consider a simple orbit of a body about the Earth where gravity is balanced against centrifugal force,

$$F_g = mg = CF = \frac{mv^2}{R} = \frac{mMG}{R^2} = \frac{Gm}{R^2}\frac{4\pi R^3\rho}{3}, \tag{12}$$

This can be solved for speed v, and the orbital period τ can then be found.

$$\tau = \frac{2\pi R}{v} = \sqrt{\frac{3\pi}{G\rho}} \sim \frac{1}{\sqrt{G\rho}} \tag{13}$$

Note from the previous discussion above that this orbital period is the same as the period of a ball falling through the earth.

3. Derivations with more appropriate Mathematics:

Fluid Expansion with GRT Lagrangian: Discussion of scalar fields like the Higgs field should use field equations. In quantum field theory, these are generally based on a Lagrangian (a concept useful for finding appropriate equations of motion based on 'least action,' A). Least Action is one of the most basic fundamental concepts in all of physics. The simplest Lagrangian in basic physics is just $L = KE - PE = mv^2/2 - PE$. A falling body on the surface of the earth, for example, will follow a parabolic path so as to give the smallest value of action $A = \int Ldt$ from its initial to its final time and location. Quantum field theory replaces kinetic energy with changes in field values:

$$\mathcal{L} = \frac{1}{2}\partial_\mu \phi g^{\mu\mu}\partial_\mu \phi - V(\phi) \simeq \frac{\dot{\phi}^2}{2} - V(\phi). \tag{14}$$

To use this Lagrangian, the principle of Least Action here says that $\delta A_\phi = \int d^4x\sqrt{-g}\delta\mathcal{L}_\phi = 0$. Notice this volume correction factor $\sqrt{-g}$ where g = determinant of $g_{\mu\nu}$. For the easy case of flat geometry , the metric is just $ds^2 = dt^2 - a^2[dr^2 + r^2(d\theta^2 + \sin^2\phi d\phi^2)]$ where 'a' is the now familiar 'scale factor' and $|g| = a^6$. For a scalar field in a spatially homogeneous universe, only the time derivative counts [5]. Then, the resulting field equation of this Lagrangian action is:

$$\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^\mu}(\sqrt{-g}\, g^{\mu\mu}\frac{\partial\phi}{\partial x^\mu}) + \frac{\partial V}{\partial \phi} = \frac{1}{a^3}\frac{\partial}{\partial t}(a^3\dot{\phi}) + V'(\phi) = \ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + V'(\phi) = 0. \tag{15}$$

Again, the middle term is $3H\dot{\phi}$ where Hubble $H = \dot{a}/a$. This kinetic 'friction' follows naturally from the universal expansion and the equations of general relativity.
For 'slow roll,' the $\ddot{\phi}$ term is negligible leaving just $3H\dot{\phi} = -V'(\phi)$, the slope of the potential, V. The fluid equation for $\dot{\phi}$ tells how fast ϕ moves away from its initially zero value.

Another approach to this key equation is to apply the following conversion equations from classical mechanics to scalar fields [6]:

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi), \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi), \qquad so\ (\rho + p) = \dot{\phi}^2. \tag{16}$$

The first terms are like a kinetic energy of the field and the second V is like a PE or binding or internal energy contributing to mass and hence density of the field. Then the Friedman equation becomes $H^2 = (8\pi G/3)[V(\phi) + \dot{\phi}^2/2]$, and the equation of motion is found from:

$$\dot{\rho} = \dot{\phi}\ddot{\phi} + V = -3H\dot{\phi}^2, \quad so\ \ddot{\phi} = -3H\dot{\phi} - \frac{dV/dt}{d\phi/dt}, \quad or$$

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \tag{17}$$

Smoot [7] refers to this equation as the 'Klein Gordon equation in a FRW Universe' analogous to the familiar form $(\Box^2 + m^2)\phi = \ddot{\phi} - \nabla^2\phi + m^2\phi = 0$. Its effective mass is given by the oscillation frequency about its minimum potential well. Adequate inflation will occur when the initial slope V' is very low and curvature of V is also very low (i.e., V high, level, and linear for some distance) [technically $\epsilon = (V'/V)^2/16\pi G \ll 1$ along with $|\eta| = V''/8\pi GV \ll 1$]. Inflation ends when $\epsilon = 1$ at the bottom of the $V(\phi)$ potential. Andrei Linde [8] calls it the 'harmonic oscillator equation' when V has the mass form $m^2\phi^2/2$. Names of current inflation models include [6]: Chaotic (standard inflation which itself contains polynomial $V(\phi)$'s, power-law, natural inflation), Multi-Field, extended general relativity, and Open inflation.

Density Perturbations:

A major goal of inflation is to account for the observed CMB temperature/density variations which later evolved into the formation of galaxies and clusters of galaxies in the present universe. During inflation, the Hubble distance has the unusual state of being small and constant while the universe expands. Another way of saying that is that the 'comoving' Hubble length decreases (that is a way of looking at a volume of the universe without taking actual expansion into account). In the usual big bang universe (BB without inflation), the Hubble distance and comoving Hubble length always increased. All scales of the universe beyond Hubble length lack causal physics (no light speed communication could connect their parts). With inflation, the scale of the universe accelerates − slowly at first and slower than light). In this epoch, much of the universe is causally connected. But prolonged acceleration and exponentially developing scale quickly take the scale beyond the physical Hubble length so that much of the universe is again causally disconnected. However, the initially causally connected regions now grow so big that the universe we can see (our Hubble length at present) is all causally connected. Since the real radiation and matter universe has a long history with deceleration, the scale again falls within the physical Hubble length. So, initially, we have scale moving outside Hubble lengths; and finally we have parts of the universe again coming back into our Hubble length: "Out and then In Again." This difference is similar to a sine wave from 0^o to perhaps 400^o and is also true in the comoving picture. So, there is an early, middle, and late stage punctuated by crossovers between cosmic scale and speed-of-light Hubble lengths.

Perturbations created in the early causal stage get "imprinted" or "frozen-in" during crossover because there is no longer any causal mechanism to change them. The mechanism creating the earliest perturbations is quantum fluctuations due to the uncertainty principle.

Cosmic Microwave Background (CMB) results measured by COBE state density perturbations near $\delta\rho/\rho = \delta_H \equiv \delta_H(k = a_o H_o) \simeq 2 \times 10^{-5}$. This is seen in the subtle hot and cold spots on the CMB microwave surface with temperature variations $\Delta T/T \sim 10^{-5}$. Different authors use different symbols for variations: Δ, δ, or even σ for standard deviations. For example, RMS surface vertical roughnesses are defined as $R = \sigma_z/|z(x)|_{average}$. In the CMB prior to last scattering, there were plasma oscillations that can be thought of as acoustic waves. Waves can be decomposed into a sum of modes with different wave numbers, $k = 2\pi/\lambda$. We see these modes in the sky, so their wavelengths are measured as angles rather than as distances. It is conventional to express density fluctuation inhomogeneities in terms of Fourier expansion

$$\frac{\delta\rho(\vec{x})}{\rho} = \frac{1}{(2\pi)^3} \int \delta_k \, e^{i\vec{k}\cdot\vec{x}} d^3k \tag{18}$$

where ρ is the appropriate mean universal density and δ_k is wavenumber amplitude. The power spectrum is merely the square of the amplitude representing wavenumbers,

$P(k) = |\delta_k|^2$ and is useful because measurements are usually made using energy rather than amplitudes. In cosmology, phases are random, so they are not needed. Roughness in cosmology grows with scale factor: $\delta\rho/\rho \propto a^2 \propto t$ for time in the radiation era, and $\delta\rho/\rho \propto a \propto t^{2/3}$ in the matter era [4]. The roughness spectrum injected by inflation is $P(k) \propto k$ for large scales. But this now rolls over for intermediate scales and falls for smaller scales $P(k) \propto 1/k^2$ or worse [4]. The literature is tricky because sometimes a volume weighted measure is used: $\Delta^2(k) \simeq (\delta\rho/\rho)_k^2 = k^3|\delta_k|^2/(2\pi^2 V)$ where V is 'the volume of the fundamental cube'[3].

The roughness today is extreme on small scales like 100 ℓy with $\Delta\rho/\rho \gg 1$, a highly nonlinear regime. The universe may appear homogeneous at the large scale, but it is highly inhomogeneous on small scales. The calculation of roughness has to be done numerically and separately for different constituent regimes [7]: scalar fields, baryons, neutrinos, radiation, DM and DE, and matter clustering. But the concern here is only about the earliest eras. During the radiation era, the rapid universal expansion prevented any gravitational clumping [4].

Inflation is able to resolve previous mysteries of the Big Bang cosmology. But it also makes predictions that can be tested. One confirmation is something called the spectral index, n_s which is equal to one for a scale-invariant spectrum [10]. The simplest models of inflation predicted n_s between 0.92 and 0.98, and WMAP spacecraft data infers that $n_a = 0.963 \pm 0.012$.

The European Space Agency's Planck satellite was launched on 14 May 2009 as the third in the series of CMB studies [12]. A group of 29 status reports came out in March, 2013 including an update on new constraints on inflation. The scalar spectral index is now $n = 0.9603 \pm 0.0073$ without any running. This rules out exact scale invariance at over 5σ; and some of the hundreds of inflation models are now ruled out. "Unless a quartic term is allowed in the potential, we find result consistent with second-order slow-roll predictions." Other results include slight changes in cosmic parameters such as $\Omega_\Lambda = 0.693$, $H_o = 68$, $Age = 13.8\, Gyr$. New detailed polarization data has not yet been published.

Paul Steinhardt was one of the developers of the "new inflationary model" but is no longer supporting it. (He also helped introduce quintessence, ekpyrotic and cyclic model alternatives to inflation). In the original Guth scenario for inflation, "the rate of expansion of the universe dominates the rate of production and growth of bubbles; the bubbles never coalesce to complete the transition" to produce a stable vacuum. The new inflation patched up this problem [Linde, Albrecht and Steinhardt, 1982]. Steinhardt now notes that observations have strengthened the 1980s version of inflationary cosmology, but the arguments against inflation have also grown stronger! [13] "The recent Planck satellite combined with earlier results eliminate a wide spectrum of more complex inflationary models and favor models with a single scalar field, as reported in the analysis of the collaboration. More importantly, though, is that all the simplest inflation models are disfavored by the data while the surviving models namely those with plateau-like potentials are problematic [14]." The vast majority of key parameters result in "bad inflation" not matching our universe. Inflation 'can only begin to smooth the universe if the universe is unexpectedly smooth to begin with!' Our flat universe is much more likely to result from starting configurations without inflation at all. The inflation of the 1980's was wrong. If inflation is 'eternal,' then it is not able to explain or predict anything! It needs 'a major fix or must be replaced. Anthropic concepts do not help, and 'a challenge for the inflationary paradigm in light of the Planck 2013 data is to explain why no significant multiverse effects have been observed. LHC data also suggests that 'the current symmetry breaking vacuum is metastable making initial conditions even more unlikely. So, there are now three major problems, "a new initial conditions problem, a worsening multiverse-unpredictability problem, and a novel kind of discrepancy between data and paradigm that we termed the unlikeliness problem."

4. Inflation and B-Mode Polarization

In March, 2014, after a three-year analysis, the South Pole telescope group BICEP2 [16] announced its discovery of what is called B-Mode Polarization (at more than 5 sigma confidence above expected background contamination). This was a difficult experiment because the B-mode signal is only about one-percent as strong as the previously analyzed weak background temperature fluctuations . The "Background Imaging of Cosmic Extragalactic Polarization" study number two (BICEP2) revealed pinwheel-like swirls in the polarization of the cosmic microwave background (CMB). The parameter expressing this curling polarization is called "r, the ratio of power in tensor to scalar density perturbations of the CMB. Its value is $r = 0.20$ and represents a much higher signal than anticipated. Over the next year, it should not be difficult for a variety of other science experiments to verify this value. This discovery is considered to be major because it may be strong evidence for the early inflationary expansion of the universe. That GUT scale blow-up has powerful gravitational waves which produce the circular polarizations. A previous belief in "Higgs Inflation is now dead because it suggests $r = 0$ instead. And the ekpyrotic model is also now untenable. What is left is a standard inflation from some new GUT field near 10^{+16} GeV as in modeling by Andre Linde. Still, further science is needed because there is a discrepancy between these Antarctic results and the present analysis from PLANCK which suggests a highest value of $r < 0.11$.

References

[1] Alan H. Guth, "Inflationary Universe: A possible solution to the horizon and flatness problems," Physical Review D, Vol. 23, No. 2, 15 January 1981, pp. 347-356.
[2] Joseph Silk, The Big Bang, Third Edition, 1980-2001, W.H. Freeman and Company, New York.
[3] Edward W. Kolb, Michael S. Turner, The Early Universe, 1990, Addison-Wesley Publishing Company.

[4] Mark Whittle, Cosmology: The History and Nature of our Universe, The Great Courses guidebook, 2008 (mathematics is given in the appendix section at the end of the guidebook). This is a truly excellent DVD course with wonderful exposition.
[5] P.J.E. Peebles, Principles of Physical Cosmology, Princeton University Press, 1993. (e.g., p 396).
[6] Andrew R. Liddle, "An Introduction to Cosmological Inflation," arXiv:astro-ph/9901124v1, 11 Jan 1999.
[7] George Smoot and Jorge Cervantes-Cota, "Cosmology today-A brief review," arXiv:1107.1789 astro-ph 9 Jul 2011.
[8] The Future of Theoretical Physics and Cosmology, Celebrating Stephen Hawking's 60th Birthday, Ed. G.W. Gibbons, E.P.S. Shellard, S.J. Rankin, Cambridge, 2003.
[9] David H. Lyth, "Which is the best inflation model?," arXiv:hep=th/0311040v1 5 Nov 2003.
[10] Inflation (Cosmology) http://en.wikipedia.org/wiki/Inflation_(cosmology)
[11] Jerome Martin, et. al., "Encyclopedia Inflationaris," ArXiv:1301.3787 [astro-ph.CO] 15 Mar 2013, 274 pages.
[12] Planck Collaboration, "Planck 2013 results. XXII. Constraints on Inflation," arXiv:1303.5082 [astro.CO] 20 Mar 2013, also see "Overview" 1303.5062 listing all XXIX reports so far.
[13] Paul Steinhardt, "The Inflation Debate," Sci Am. April, 2011, e.g., see: http://www.physics.princeton.edu/ steinh/0411036.pdf
[14] Anna Ijjas, et. al., "Inflationary paradigm in trouble after Planck2013," ArXiv:1304.2785 [astro-ph.CO] 9 Apr 2013.
[15] Andrei Linde, "Inflationary Cosmology," ArXiv:0705.0164 2 May 2007, 45 pgs.
[16] BICEP2 http://bicepkeck.org/ BICEP Release package. ArXiv: 1403.3985 BICEP2: "Detection of B-mode Polarization at degree angular scales." http://arxiv.org/pdf/1403.4302.pdf "BICEP2: II: Experiment and Three-Year Data Set," Mar 17,2014, astro-ph. Also see Planck 1403.3919.

HAWKING RADIATION

DP

Abstract. Popular books on Hawking radiation (e.g., Susskind [1]) provide elementary perspectives but leave many basic questions unanswered. It is the goal of this note to give a slightly deeper view for mathematically inclined amateurs. There are some simplified derivations approximating the results for Hawking entropy and temperature that are more intuitive than those depending so much on the details and art of general relativity and quantum field theory. This note discusses the history and algebra of the topic but avoids the issue of preservation of information via quantum entanglements at the horizon. Also included is a formula for surface temperature that seems hard to find elsewhere.

1. Introduction

In classical thermodynamics, there are four basic macroscopic laws:
(Zeroth Law) Systems in thermal equilibrium have the same constant absolute temperature, T, throughout.
(First Law) The total energy of an isolated system is always conserved.
(Second Law) The entropy, S, of a thermally isolated system cannot decrease [$dS/dt \geq 0$, and if not isolated then $d(Heat) = TdS$] [1], and
(Third Law) The entropy of a system has a limiting property: $\lim_{T\to\circ} S = S_o$, and it is impossible to achieve $T = 0$ by a physical process.

Separately, the classical laws of black hole mechanics could be stated as [2]:
(0) The horizon of a stationary black hole has constant surface gravity, g.
(1) For perturbations of stationary black holes, the incremental change of energy is related to change of area, $dM = dE/c^2 = |g|dA/8\pi G$ [ignoring possible changes in angular momentum, dJ, and charge, dQ].
(2) The horizon area, A, is a nondecreasing function of time, $dA/dt \geq 0$.
(3) It is not possible to form a black hole with vanishing surface gravity, $g = 0$.

There is some apparent similarity between these two sets of laws if we approximate the surface gravity of a black hole with temperature and the area of the event horizon with entropy ($T \sim |g|$, $S \sim A$ – at least up to some multiplicative constants). And, of course, energy (or heat) maps to mass via $E = Mc^2$. Jacob Bekenstein may have been the first

email: davepeterson137@gmail.com.

For Cosmology Club. October 23, 2013, Last update Nov. 18, 2013, dp.

[1]Sometimes, $dE = TdS+$ work terms is included under the First Law.

to notice this analogy and publish comments on it in 1972 [3] [2]. He thought that black holes have maximum entropy above anything else over the same volume of space. Stephen Hawking thought that Bekenstein had carried the analogy too far and wished to disprove it − but he instead finally verified the analogy as being real. Black holes have entropy and they must radiate heat. With his black hole temperature and evaporation, he also added a new temperature attribute to black holes and showed that mass, area, and surface gravity could decrease when quantum effects were considered along with the classical effects.

This raised important questions about the role of information in black holes. The earlier "No-Hair Theorem" (actually a conjecture) said that regardless of what falls into a black hole, the result is completely characterized simply by the black hole mass, spin, and possible net charge (M, J, Q). That implied that all other incoming information was at least invisible to the external world. In 1974, Hawking showed that black holes may also evaporate [4] so that any information that may have been trapped inside or on the surface was completely destroyed when the black hole vanished along with its central singularity. While largely accepted by general relativists, this loss of information annoyed some quantum field theorists (i.e., particle physicists). Many publications attempted to clarify this issue with some apparent successes but ultimate confusion (as of 2013, e.g., the "AMPS firewall" reference [12]). In 2004 and subsequently, Hawking himself finally accepted the preservation of information as existing in the outgoing Hawking radiation [8]. Part of the research that finally persuaded him was a new esoteric result called the "AdS/CFT" duality conjecture between "supergravity in anti-deSitter space and a conformal field theory" on its boundary [11]. Whether that is really relevant won't be discussed here.

This arena is hard to comprehend because its details require an understanding of general relativity and quantum field theory together. One difficult aspect is that the view of what is real depends on who is doing the observing. The Schwarzschild radius is derived as an external matter free solution of the Einstein equations. At that radius, a very different interior solution begins. The transition between these regions became clear with the use of what are called Kruskal coordinates in 1960 [7]. There were some hints before this that the Schwarzschild 'singularity' might not be "real" but merely an apparent singularity due to a choice of coordinates. A freely falling observer would pass right through the radius R without noticing it (unless there is really something to the newly introduced "firewall" concept).

In 1974, Stephen Hawking wrote a paper in the journal 'Nature' entitled "Black Hole Explosions" [4] in which he proposed that black holes radiate with a black body temperature and can ultimately evaporate explosively. Any information falling through the black hole horizon will be destroyed by falling into the inner singularity which eventually vanishes. We wish to address some of this modestly in the next section.

[2]1972 is also the year he obtained his Ph.D. from Princeton University, and his Ph.D. thesis said that black hole entropy is identified with its surface area.

2. A Little Basic Mathematics

The traditional discussion of black holes begins with the Schwarzschild metric solution of the Einstein field equations outside a highly condensed spherical mass. The word 'metric' is a way of measuring distances over small displacements in the coordinates being used. For our usual Euclidean space with three dimensions (E^3), the Pythagorean theorem gives distances as: $ds^2 = dx^2 + dy^2 + dz^2$. An overall form for a 3D metric is $ds^2 = g_{ij}dx^i dx^j$ with coordinates labeled dx^1, dx^2, dx^3 and superscripts just standing for coordinate 1, 2 and 3. 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 $g_{i\neq j} = 0$.

For <u>special relativity</u>, the reference is **light** with speed c; and the metric this time is the difference between time and space increments: $ds^2 = c^2dt^2 - dx^2$. A 'timelike' 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, $dx/dt = c, so ds^2 = 0$. For a particle with mass and $v < c$, we can write $d\tau^2 = dt^2 - dx^2/c^2 = dt^2(1 - (dx/dt)^2/c^2) = dt^2(1 - (v/c)^2)$. Then $dt = \gamma d\tau = d\tau/\sqrt{1 - v^2/c^2}$. The 'Lorentz factor' $\gamma \geq 1$; so perceived time duration is larger than the clock time in the frame of the moving object. Then a muon 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.

For <u>general relativity</u>, the metric coefficients represent curved space and time and become functions instead of just numbers.

The 'Robertson-Walker' metric for cosmology still has $g_{oo} = 1$ for time, but space can have negative curvature, positive curvature (like the 3-sphere S^3), or zero flat space (like E^3). But space also expands with time.

For a ball of mass in space, recall that the non-rotating Schwarzschild exterior solution to the Einstein general relativity equations (GR, $R_{\mu\nu} = 0$) has metric coefficients of the form

(1)
$$ds^2 = c^2d\tau^2 = g_{\mu\mu}(dx^\mu)^2 = \left(1 - \frac{2MG}{c^2r}\right)(cdt)^2 - \frac{dr^2}{(1 - 2MG/c^2r)} - r^2(d\theta^2 + \sin^2\theta(d\phi)^2)$$

which has an apparent singularity ('blows-up') at the **Schwarzschild radius**,
$\boldsymbol{R = r_{Schwarz} = 2MG/c^2}$, where M is the mass of a compact spherical body.
The index mu runs from $\mu = 0 = time$ and metric coefficient g_{oo} represents time curvature. Then $\mu = 1 = r$, $\mu = 2 = \theta$, and $\mu = 3 = angle\ \phi$. Since $E = Mc^2$, and $M = Rc^2/G$, the total energy of a black hole can be written as: $E = Rc^4/G-$ energy is proportional to R.

Schwarzschild coordinates use a radial coordinate, r, chosen so that angular measures ($r\Delta\theta$, $r\sin\theta\Delta\phi$) appear Euclidean: a circle has the usual circumference $C = 2\pi r$ and a 2-sphere has area $A = 4\pi r^2$. All of our 'units' systems are based on our own size, weight, temperature, and human speeds; but physical Nature sees things on its own terms. So, many articles on GR use 'natural' units where one sets basic natural constants to unity: $G = c = \hbar = k = 1$ (k is Boltzmann's thermal constant). Then $R = 2m$ with little $m = GM/c^2$ thus hiding the G and the c constants (but, for our own comfort, we'll usually include those values here as if we were using the MKS system ['System International,' SI]). For black holes with mass in terms of the mass of our sun, $R \sim 3\,km\,(M/M_{\odot})$. The value of R is called the 'Event Horizon.' At that radius, time flow becomes

$$\frac{dt}{d\tau} = \sqrt{{g_{oo}}^{-1}} = \frac{1}{\sqrt{1 - 2MG/c^2 r}} \to \infty \quad as\, r \to R\, from\, outside. \tag{2}$$

The variable 't' is coordinate time at a distant observer, and tau τ is 'proper time' local to a body being studied [3]. If the sign of g_{oo} is positive (i.e., metric signature $+---$), then the metric distance $ds = cd\tau$. This equation means that observed time flow slows to a stop at the horizon, R. That is one reason that black holes used to be called 'Frozen Stars' by Soviet physicists (but ultimately, 'Black Holes' by John Wheeler in 1967). Because observations diminish and vanish at $r = R$, Wolfgang Rindler gave it the new name, "horizon."

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 $\pm 1\%$ for the 'Pound-Rebka' experiment over $\Delta h = 22.5\,m$ back in 1959 [16]] [4]. For black holes, red-shifting is much more profound. Frozen time flow means long observed period means perceived very low frequency or huge red shifting. So Hawking radiation experiences very strong red shifting when moving away from the horizon. That also means that the near-horizon must be very hot $-$ at least from the perspective of an observer held above the horizon 'by a rope' (which is equivalent to a powerful acceleration from a rocket ship against the force of gravity).

To avoid infinite changes and blow-ups at the event horizon, it is a convention to refer instead to Kip Thorne's "membrane" formalism as a surrogate for the black hole [14]. Thorne doesn't specify an exact location for this spherical surface, just pick one up close but still mostly outside the "frozen boundary layer." Similarly, a "stretched-horizon" [9] one Planck length above the event horizon enables the discussion of particular huge but still finite numbers there. The temperature on the stretched horizon (before red-shifting down to much lower Hawking temperature. T_H) is the Planck temperature $T_P \sim 10^{32}$ K. In general, $T_H(far) = \sqrt{g_{oo}}\; T_{up\,close}$.

[3] Consideration of the dr (or g_{rr}) term would contribute a velocity dependent Doppler shifting.

[4] Actually, weak field red-shift can be derived without General Relativity by simply using the principle of equivalence and special relativity (see Schiff [18]).

The gravitational acceleration at the Schwarzschild radius is simply given by:

$$|g| = \frac{GM}{R^2} = \frac{GM\,c^4}{(2GM)^2} = \frac{c^4}{4GM} \tag{3}$$

Using this equation for g, one could express Mass as $M = gA/4\pi G$.

Hawking came up with an effective black body temperature for black holes now called the Hawking temperature:

$$T = T_{Hawking} = \frac{hc^3}{16\pi^2 GMk} = \frac{\hbar c^3}{8\pi GMk} = \frac{c^4\,\hbar}{4GM\;2\pi kc} = \frac{|g|\hbar}{2\pi kc}. \tag{4}$$

Particles are produced with zero net energy in the curved spacetime close to the event horizon of a black hole. This claim and formula was initially met with much skepticism but later with general acceptance as a variety of independent approaches all agreed with it. Note that adding mass/energy to the black hole causes a lower temperature. That represents a negative specific heat; and that is typical for self gravitating systems.

Note that at peak energy of black body radiation, $h\nu_{max} = hc/\lambda \sim 3kT$ (Wien's law). Then using $T_{Hawking} = hc/8\pi^2 kR$, we get $\lambda \sim hc/3kT \sim R8\pi^2/3 \sim 8\pi R$. So, peak wavelength is crudely approximated by the Schwarzschild radius.

In 1976, Bill Unruh [5] discovered what is now called the "Unruh effect," which predicted that an accelerating observer will observe black-body radiation where an inertial observer would not. Accelerating detectors find themselves in a warm background with a temperature proportional to the acceleration, a. That is, $T = \hbar a/2\pi ck$. This looks like $T_{Hawking}$ but with an 'a' instead of a $|g|$. The Unruh effect is an analog of Hawking's effect for the case of accelerating frames in flat space, but it can be considered to be more fundamental than the Hawking effect. The distance from the accelerating detector to its effective horizon is comparable to the gravitational Schwarzschild radius, R.

There is simple heuristic derivation of the effect (e.g., [6]) [5]: Using an acceleration 'a' over a distance Δx, consider the creation of an electron-positron pair with needed energy $\Delta E = 2mc^2 = (ma)\Delta x$. Now use the uncertainty principle in the time-energy form: $\hbar/2 \sim \Delta E \Delta t \sim \Delta E(\Delta x/c)$ [6]. So $\Delta E = \hbar c/2\Delta x = \hbar a/4c$. Then if thermal agitation energy for a single electron is $E = 3kT/2$,[7] we have:

$$T = \frac{\hbar a}{6ck} \sim \frac{\hbar a}{2\pi ck}. \tag{5}$$

[5] Not a detailed computation (which is hard); just ideas for approximation.

[6] Technically, time is not a dynamical variable, and there is no operator for time. The operator for energy is the Hamiltonian.

[7] The '3' comes from 3-degrees of freedom (x,y,z). Or alternately, one photon up and one photon down and use Wien's law mean photon energy $E \sim 3kT$ for assumed black body radiation [17]. That is, peak frequency is $\nu_{max} = bT = 58.8\,GHz\,T$, and we want $h\nu = nkT = hbT$, $so\,n = hb/k = 2.82 \sim 3$.

which has nearly the same form as for Hawking temperature if the gravitational field at the horizon is replaced by the acceleration, a.

Notice the assumptions that went into this. One is that the quantum vacuum is always having fluctuating particle production such as virtual electron-positron pairs (Quantum Field Theory, QFT, is assumed). Energy is force times distance, $Fd = mad$, so if mass blips into momentary existence and an accelerating field is present, more energy can be pumped into the mass. Initially, creation energy is borrowed temporarily from the vacuum, but more energy may make it more permanent. A detector accelerating is similar to cosmic fluid accelerating with respect to the detector. In cosmology, the usual Cosmic Horizon (or 'light horizon' or 'particle horizon' now) is at a distance where the recession velocity is the speed of light (about 16 billion ly). No objects beyond that can be presently communicated to us. At the ultimate speed limit, $v = c$, the Newtonian ('gun barrel') formula $v^2 = 2ad \rightarrow c^2$ implies that at a distance $d = c^2/2a$ from the detector, there will be an effective horizon even in flat space. This value happens to also be the gravitational Schwarzschild radius, d = R.

From here, we can discuss black hole entropy, S_{BH}, using an elementary derivation stated in Susskind [1]:

The initial trick used by Susskind (and previously by Jacob Beckenstein) was to only consider radiation wavelengths similar to the size of a black hole (so $\lambda \sim R_{Schwarzschild}$). This was later justified by stating that although an infalling observer may have very localized information at the event horizon, distant external observer will think the information is spread out uniformly over the entire stretched horizon, H_s, before being re-radiated. The goal of using H_s is to avoid any reference to events inside a black hole. The de Broglie wavelength of a photon having momentum p is $p = h/\lambda$, so $E_{photon} = pc = h\nu = hc/\lambda \sim hc/R = \Delta Mc^2$. If a photon is emitted by a black hole, the mass will slightly decrease by an energy $\Delta E = \Delta Mc^2$ and the black hole radius would decrease slightly by a value ΔR. It will turn out that information bits can be represented by tiny Planck size areas on the surface or horizon of the black hole. So, we also need to know this size: $\ell = \ell_{Planck} \equiv \sqrt{\hbar G/c^3} \sim 1.6 \times 10^{-35}\ meters$. Now the altered radius is:

$$R + dR = \frac{2MG}{c^2} + \frac{2G\Delta M}{c^2},\ or\ dR = \frac{2hG}{Rc^3} = \frac{4\pi\hbar G}{Rc^2}. \tag{6}$$

$$Area = A = 4\pi R^2, \Rightarrow dA = 8\pi R(dR) = 8\pi R\left(\frac{4\pi\hbar G}{Rc^3}\right) = 32\pi^2\ell^2 \quad \propto \ell^2. \tag{7}$$

The increase in area due to a new bit of information is proportional to the 'Planck area.'

Temperature is the increase in the energy of a system when you add one bit of entropy, $T = dE/dS$, so $dE = TdS$. Now apply Hawking's claim that $T = \hbar c^3/8\pi GMk$ from equation (4) so that:

$$dE = c^2 dM = TdS = \frac{\hbar c^3\, dS}{8\pi GMk},\ so\ \ \frac{dS}{k} = \frac{8\pi GM(dE = c^2 dM)}{\hbar c^3} = \frac{4\pi d(M^2)G}{\hbar c^3}. \tag{8}$$

Now integrate this [with initial condition S(M=0) = 0], and use $R^2 = (2MG/c^2)^2$ to get:

$$\frac{S}{k} = \frac{4\pi R^2 c^3}{4G\hbar} = \frac{A\ c^3}{4\,G\hbar} = \frac{A}{4\,\ell^2}. \tag{9}$$

Numerically, the total entropy of a black hole is proportional to R^2 and is also one-fourth of its surface area broken into tiny Planck areas [8]. Historically, Jacob Bekenstein proposed the idea that a black hole had entropy proportional to its surface area divided by Planck area in 1973 [3]. But it was Hawking who supplied the mechanism and calculated the factor "one-fourth" in 1974 [4]. This equation (9) for entropy is referred to as the "Bekenstein-Hawking" formula. It is also used as a maximum "Holographic Bound."

Combining the last two equations gives:

$$dM = \frac{TdS}{c^2} = \frac{|g|\hbar \quad c^3 k\, dA}{2\pi k c^3\ \ 4G\hbar} = \frac{|g|\, dA}{8\pi G}. \tag{10}$$

And this is the form appearing on page one as the First Law of black hole mechanics. A change in mass or energy corresponds to a change in the surface area of the horizon.

Black hole articles do not seem to discuss the details of what information is except to say that it is measured by entropy. That is, entropy is information − or more precisely "lack of information" (Shannon, 1948). A stress on the concept of unitarity seems to imply that the information is that contained in wavefunctions as opposed to classical information directly involving mass, charge, angular momentum, spins, or particle numbers. Perhaps Bekenstein-Hawking entropy is the "entanglement entropy" between the interior and exterior regions of a black hole (an entanglement measure for bipartite pure states). An example for pairs of horizontal versus vertical polarization entangled photons used in the first experimental verification of quantum teleportation [22] is the spatially antisymmetric wavefunction $|\Psi^-\rangle = (|H\rangle|V\rangle - |V\rangle|H\rangle)/\sqrt{2}$ (information about polarization superpositions and phase relationships for photon 1 and photon 2 going off in different directions).[9]

3. Some Numbers

For a better perspective, let us consider two special cases: a lowest mass black hole near 4 suns ($M = 4M_\odot$) and one a million times bigger like the newly established super black hole "Sgr A^* " in the constellation Sagittarius in the middle of our Milky Way galaxy (MW). The approximately known mass of this 'SBH' is $M = 4 \times 10^6 M_\odot$. The mass of our sun is $M_\odot = 1.99 \times 10^{30}\ kg$ with a radius of $R_\odot = 6.96 \times 10^8$ meters. The surface gravity

[8]The middle form, $S = Akc^3/4\hbar G$: On the occasion of his 60th birthday ([21] p. 113) Hawking said, "I would like this simple formula to be on my tombstone."

[9]Perhaps more relevant at the horizon would be an entangled pair given by:
$|\Psi\rangle_{pair} = (|0\rangle_{in}|0\rangle_{out} + |1\rangle_{in}|1\rangle_{out})/\sqrt{2}$ [25].

of our sun is $g_\odot = 274\, m/s^2$, but for black holes it will usually be much much larger. The entropy of the sun is $S_\odot \sim 10^{58}\, k_B$.

Calculation for the 4 sun BH gives a radius of $R = 12$ km, and entropy is $S = 10^{78}$ times Boltzmann's constant, k, (which we might set equal to unity). This BH entropy is a great deal higher than that of a star from which it came. For practical purposes, information does disappear in the presence of black holes, and the horizon is a proxy for the entropy that has gone beyond the horizon (Bekenstein). It is huge because of all of the incoming information that is lost. In 1961, IBM's Rolf Landauer made the claim that information is physical, and any erasing of information produces heat [13] [10]. The Hawking temperature is $T = 16$ nano-Kelvins, and surface gravity is $g = 400$ billion earth g's [tiny temperatures and crushing surface gravities].

The super BH for the MW gives a radius of $R \sim 8$ sun diameters, $S = 10^{90}\, k_B$, $T = 16\, femto \cdot kelvins$, $g = 400$ thousand earth g's. Density is mass over volume using $V = 4\pi R^3/3$, or $\rho = 3c^6/32\pi G^3 M^2$. So, the density of the $SgrA^*$ BH turns out to be the same as for <u>water!</u> But the $4M_\odot$ BG has a density $10^{12}\times$ larger.

There are even bigger supermassive black holes in other galaxies. The SBH in the nearby Andromeda galaxy ($M31^*$) may weigh roughly 200 million solar masses [30]. M87 in the Virgo cluster has a 6.4 billion sun mass (and the Hubble telescope shows a long one-sided jet from the core extending about 5000 ℓyr away). And NGC 4889 (maybe a dead quasar) in the Coma cluster has a mass of $\sim 20\, giga\, M_\odot$! These have larger sizes and lower densities. For a 20 billion suns SBH, surface gravity is 763 m/s^2 (about three times stronger than that of our sun), size is about ten times larger than our solar system, and density is only 4 percent that of <u>air</u> in our atmosphere. It is difficult to think of a nebulous transition from vacuum to air as being such a significant place as to have a "firewall."

4. Discussion

The available literature contains a barrage of black hole formulas that can largely be deduced from the equations above. They like to continue the use of Planck scale formulas using for example Planck mass, $M_p = \ell_p c^2/G = \sqrt{\hbar c/G} \sim 1.2 \times 10^{19}\, GeV/c^2 = 2.2 \times 10^{-5}\, grams$ [11], and Planck temperature,
$T_p = M_p c^2/k_B = \sqrt{\hbar c^5/G k_B{}^2} \sim 10^{32}$ K. Since natural units set $G = c = \hbar = k_B = 1$, then all the Planck units are just unity (1). If they say, e.g., $T_H \sim M_p{}^2/M$, they mean that Hawking temperature is $T_H = (M_p{}^2/M)(c^2/8\pi k)$; and they may write this as just

[10] "Landauer's Erasure Principle states that the erasure of 1 bit of information requires a minimum energy cost equal to $kT\, ln(2)$ where T is the temperature of a thermal reservoir used in the process". Its acceptance has grown slowly over the past five decades.

[11] So, the Planck distance $\ell_p = M_p G/c^2$ resembles a gravitational radius of a Planck mass.

$T_H = 1/8\pi M$. The BH entropy may be written as $S_{BH} = 4\pi M^2 = A/4 = 1/16\pi {T_{BH}}^2$.

One of the most important observations is what happens to the metric coefficient $g_{oo} = (d\tau/dt)^2 = (1 - 2MG/c^2 r)$ from our first two metric equations (1 and 2) above. On the stretched horizon, H_s [9], we wish to set the coordinate radius r to $R_{Schwarschild} + \ell_p$ where $R \gg \ell$. And the purpose of this H_s horizon again is to work with real numbers instead of infinities and zeros. So:

$$\text{(11)} \qquad \frac{d\tau^2}{dt^2} = g_{oo}(r = R + \ell) = 1 - \frac{R}{R + \ell} \cong \frac{\ell}{R} = \frac{\sqrt{\hbar G/c^3}}{2MG/c^2} = \frac{\sqrt{\hbar c/G}}{2M} = \frac{M_p}{2M}.$$

This is a very small real number instead of being zero on the Event horizon.

Now, using the formula for the Planck temperature in a paragraph above, M_p is the same as kT_p/c^2 (or we might just say, $M_p \sim T_p$). Look again at one of the expressions for Hawking temperature:

$$\text{(12)} \qquad T_H = \frac{{M_p}^2\ c^2}{M\ \ 8\pi k} = \frac{(kT_p/c^2) M_p\ c^2}{M\ 8\pi k} = \frac{T_p\ M_p}{8\pi\ M} = \frac{T_p\ g_{oo\ stretch}}{4\pi}$$

Since T_H is seen at $g_{oo}(r \to \infty)$, this might suggest that the temperature on the stretched horizon, H_s is near a claim of super hot T_p. This is counterintuitive since one would expect energies (and temperatures) to transform as $d\tau/dt = \sqrt{|g_{oo}|} = \alpha$ and obey strict gravitational red-shifting like frequencies and time.[12] This function is so important that it is given a special name of α = 'lapse-function' or 'gravitational red-shift function' [14]. In addition, Kip Thorne transforms temperatures using $\sqrt{g_{oo}}$ [14] [16]. This knowledge dates at least back to 1934 (Tolman), and black body temperature should vary with photon frequency which also depends on $\sqrt{g_{oo}}$.

Surface Temperature of a Black Hole: So, what is the resolution of the problem given by equation (12)? Clearly, the temperature on the stretched horizon cannot be the Planck temperature near $10^{32}\ K$. A discussion in 'Notes' at end shows that the temperature on the stretched surface is:

$$\text{(13)} \qquad T_\ell = T_H \sqrt{g_{oo}(r = R + \ell)} = T_H \sqrt{\frac{2M}{M_p}} = \frac{T_p}{4\pi c}\sqrt{\frac{kT_p}{2M}}$$

Then, the smallest black hole at $M \sim 4M_\odot$ should have $T_\ell = 4.3 \times 10^{11}\ K \ll T_p$.
So, what is the relevance of the Planck Temperature, T_p? It applies to the final stages of black hole evaporation [28]. At that tiniest stage, the Hawking temperature is near the Planck energy.

Going one step further, since the Hawking temperature of this 4-sun black hole is only 16 nano-kelvins, the actual 3 degree K cosmic black body **(CMB)** radiation background presence is 175 million times larger. This will blue-shift down to the horizon to a much

[12] I think Susskind [9] (pg. 38 eqn. 5.1) made a mistake in his black hole surface temperature discussion. He states $T_S \sim M_p$ in eqn. 5.3 of the ArXiv version prior to PRD.

hotter temperature near $T_s \sim 7 \times 10^{19}\ K$. These infalling photons will add to the mass of the black hole and overwhelm its negligible Hawking radiation loss. The black hole will not evaporate.

Understanding the behavior of flat space Unruh frames and the flat approximation to space close to the Schwarzschild horizon is facilitated by using flat (Minkowski like) 'Rindler' coordinates (see 'Notes' at end). These coordinate systems describe 'a uniformly accelerating frame of reference in Minkowski space.' The Schwarzschild surface gravity, $g_H = 1/4M \ \ (really \ = c^4/4GM)$ suggests an accelerated frame with metric: $ds^2 = [(g_H)z]^2dt^2 + dz^2$ [14] where z is an altitude above the BH horizon, and g_{oo} looks like $[z/4M]^2$ is time-lapse-squared. Because of the principle of equivalence, all causal horizons (including deSitter horizons) approximate a Rindler horizon over a sufficiently small region.

It is unlikely that the gravitational Hawking effect nor the Unruh acceleration effect will ever be subject to measurement. However, there is an alternative system with the same formula for temperature, this time for phonons [15]. That is:

"... as pointed out by William Unruh in 1981, there exist physical systems which display a profound analogy with the Hawking radiation and which are susceptible to be observed in the lab. One of these consists of sound waves traveling in an accelerating fluid that flows across a bottleneck where it reaches supersonic velocity. Sound waves propagating against the flow may row up the stream where the fluid velocity is subsonic, but they will be dragged down stream where the velocity is supersonic."

Physicists are inventive and may come up with another alternative test such as, electron acceleration "by a standing wave formed by two counter-propagating, ultra-intense laser pulses."

Although Hawking's results are often viewed as introducing perspective on quantum gravity, the effects did not derive from Einstein's equations but are instead just a consequence of the existence of horizons (like the Unruh effect). Hawking worked from a perspective of a distant observer, but his effect can be treated as local phenomenon near horizons [10].

Using some of the black hole horizon ideas above, some physicists have extended them to the universe as well. Perhaps the total energy in any region can't be larger than a black hole of that size. If the region is the Hubble radius, $r_h \sim 1/H$, then vacuum energy would be bounded by $\rho_\Lambda \sim {M_P}^2 H^2$ (and see 'Notes' at end). The Holographic principle states that the world can be understood as a hologram encoded on a boundary to a region such as the horizon. So,

$$\rho_{space} \sim \frac{E = Rc^4/G}{Vol = 4\pi R^3/3} = \frac{3c^4 R}{4\pi G R^3} = \frac{{M_p}^2 3c^3}{R^2 4\pi\hbar} \sim \frac{{M_p}^2}{R^2}. \tag{14}$$

Suppose we try the light-horizon of the Universe at about 16 billion light years for our universe. So $R = 16 \times 10^9 \times 3.156 \times 10^7\, s\, c = 5.05 \times 10^{17}\, light\ seconds \times c$. $\hbar = 1.054 \times 10^{-34} Js$, and $M_p = 2.2\times10^{-8}kg$. So, density is about $\rho_{space} \sim 1.29\times10^{-9}J/m^3 \sim 1.4\times10^{-26}kg/m^3$. And this approximately matches the critical density of the universe of roughly $\rho_{crit} = 3{H_o}^2/8\pi G \sim 10^{-26}\ kg/m^3$! (see notes at end). So, is this a statement about dark energy from a holographic point of view? Or is it merely a statement that our universe is similar to a universal size black hole?...

String Theorists have also been contributing to the Hawking Information Paradox and have some interesting new ideas:
FUZZBALLS: String theorists believe the information paradox is solved by replacing classical black holes with a ball of fundamental strings everywhere inside the event horizon [27]. The surface is slightly fuzzy or misty which leads to the name, 'fuzzballs.' At the stringy surface of the fuzzball, the escape velocity is still that of the speed of light. Objects falling into a fuzzball get absorbed into the surface, and information gets distributed in the interior. We know classically that black holes get less dense as their mass increases (or equivalently, as the number, N, of quanta going into them increases). For fuzzballs, this is due to a growth of string length and an effective loss of string tension due to strings fusing together into larger more complex strings having distributed or fractional tension. This also means that the size of a string-ball can become quite large, and calculation agrees with the classical black hole size! [26]. However, microstates are technically now horizonless and singularity-free. Size varies as some power of N. In turn, this means that the effects of quantum gravity are not restricted to the tiny Planck length but can rather extend to huge sizes! So, some key assumptions that went into Hawking's work are no longer honored, and conservation of information is now allowed.

In classically described black holes, Hawking's argument is sound, and unitarity has to be violated (information not conserved). It is a mistake to believe that AdS/CFT duality solves the information problem, new physics is needed. In Samir Mathur's fuzzball complementarity theory [25], no quantum information gets squashed out of existence, and Hawking radiation can be unitary. The AMPS 'firewall' is based on classical black hole complementarity rather than fuzzball complementarity. In the new proposed complementarity, freely falling energy in the Hawking energy range cannot experience free fall at the horizon. But much higher energy infall can pass right on through. In fuzzball complementarity, spacetime ends in string sources outside $R = 2m$ without a horizon. This surface emits unitary Hawking radiation.

5. Personal Opinions

OK, so I'm not an expert in this arena, but I still have some opinions about it.

One, of course, is that it is very dangerous to glibly discuss physics at energies 18 orders of magnitude (powers of ten) beyond the state of experimental art. Speculation over a few extended orders is allowed and encouraged; but 18? Does physics really exist at the Planck scale. I would hope that ultimate reality exists and converges well before this scale. Also, I expect that at or near the Planck scale, physics should be essentially a 'theory of everything' (TOE). String theory (why is it called a theory if it cannot predict or be measured?) simply assumes quantum mechanics. I would hope that a TOE would 'explain' quantum mechanics instead of just assuming it. Even ordinary laser-lab experiments on quantum mechanics and entanglements are still very mysterious. It may someday be possible to show that string theory has achieved 'coherence' or overall self-consistency, but for science it also has to correspond to measured reality.

To avoid talking about infinite changes at the black hole horizon, a "stretched-horizon" in introduced one Planck-length above the Event horizon. This may be an arbitrary convention with the assumption that Planck dimensions represent some sort of 'rock bottom' or ultimately extreme limiting values (which may or may not be true).

It is assumed that quantum mechanics operates 'all the way up' and 'all the way down.' Black hole information preservation assumes a huge extension of S-matrix unitarity preserving quantum information. The entanglement mechanisms are becoming pretty involved. The Copenhagen Interpretation is trying to preserve itself by now adding on the idea of decoherence near macro-measuring devices. Instead of wavefunction collapse (and destruction of quantum information due to measurement), it is now suspected that there is only an 'appearance' of collapse due to decoherence. But there are a great many interpretations of quantum mechanics, and a large number of people would disagree. Consensus does not exist.

Despite being the King of the Sciences, there is still some fad and fashion in physics. It is subject to bandwagon effects, economics, heirarchy and sociology. For example, there are over 50,000 publications on supersymmetry, despite no experimental hint of its real existence. Some speculation is good, but this sounds like an entrenched industry of researchers may have gone overboard. The same may apply for string theory. It is natural to follow the smartest guy in the room (e.g., Ed Witten) and hope that movements approach a proper 'reality.' But there are many counterexamples (e.g, McNamara, Kissinger). And Wolfgang Pauli prevented some younger researchers from getting the Nobel prize because they listened to him when he was wrong (the spin of the electron went through against Pauli's advice just due to serendipity). We must beware of an untestable realm of physics that could be forever (horrors) "faith based."

It is assumed that the Vacuum is constantly in a state of violent fluctuation − for example always producing virtual electron-positron pairs and fluctuations of all the other fields. But, since calculations of the cosmological constant based on this QFT assumption fail drastically; it should be suspect. Kip Thorne's claim about this is that particle creations are borrowed from the Vacuum so that negative borrowing and positive emerging may cancel out for zero net vacuum energy. Perhaps these fluctuations only occur near strong fields or in the presence of other matter. For example, the Lamb Shift of hydrogen occurs in the presence of a powerful electric field from a nucleus. Perhaps this doesn't happen much in otherwise 'empty space.' Perhaps the gravitational field at the black hole horizon acts differently from strong electric fields and really doesn't produce Hawking radiation after all.

References

[1] Leonard Susskind, The Black Hole War (My battle with Stephen Hawking to make the world save for Quantum Mechanics), Little, Brown and Co., N.Y., 2008.

[2] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," Comm. Math. Phys. Volume 31, Number 2 (1973), 161−170.

[3] J. D. Bekenstein, "Black holes and entropy," Phys. Rev. D 7:2333−2346 (1973). Also see: J. D. Bekenstein, "Generalized second law of thermodynamics in black hole physics," Phys. Rev. D 9:3292−3300 (1974). And J. D. Bekenstein, "Statistical black hole thermodynamics," Phys. Rev. D12:3077- (1975).

[4] Hawking, S. W. (1974) "Black hole explosions?" Nature 248 (5443). This was followed by: Hawking, Stephen W. (1975)."Particle creation by black holes". Communications in Mathematical Physics 43 (3): 199−220.

[5] W. G. Unruh. "Notes on black hole evaporation." Phys. Rev., D14:870, 1976. [Followed by "Origin of the Particles in Black-Hole Evaporation," Phys. Rev., D15, 365-369, 1977].

[6] Michael R.R. Good, "Thermalizing the Vacuum," Dec 3, 2006, http://www.unc.edu/ mgood /research/Unruh.pdf

[7] Kruskal, M. D., 1960, "Maximal extension of Schwarzschild metric," Phys. Rev. 119, 1743-1745. (With prior hints by Eddington (1924), Lemaitre (1933), Synge (1950), Fronsdal (1959) and finally Szekeres (1960)).

[8] S. W. Hawking, "Information Loss in Black Holes," arXiv:hep-th/0507171 15 Sep 2005. This is better described by John Baez's blog, "Hawking pays off bet," for the 7/25/04 GR17 conference in Dubllin,

[9] Leonard Susskind, et. al, 28 Jun 1993. "The Stretched Horizon and Black Hole Complementarity," ArXiv.org hep-th/ 9306069. And, Physical Review D 1993 Vol: 48(8):3743. DOI: 10.1103/physrevd.48.3743 (published 15 Oct 1993).

[10] George F R Ellis, "Astrophysical black holes may radiate but they do not evaporate," ArXiv 1310.4771 gr-qc 17 Oct 2013. Also see follow up on 1310.0595.

[11] J.M. Maldacena, "The Large N Limit of Superconformal Field Theory and Supergravity," Adv. Theor. Math. Phys. 2, 231 (1998) and also ArXiv:hep-th/9711200

[12] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully, "Black Holes: Complementarity or Firewalls?", ArXiv:hep-th 13 Apr 2013. [AMPS].

[13] Landauer, R., "Irreversibility and Heat Generation in the Computing Process." IBM J. Res. Dev. 1961, 5, 183-191. In addition, see: Sean Carroll, "Thanksgiving for Landauers Principle," 11/28/13, http://www.preposterousuniverse.com/blog/2013/11/28/thanksgiving-8/ and for an opposed view, http://www.pitt.edu/∼ jdnorton/papers/Eaters.pdf.

[14] Kip S. Thorne, Richard H. Price, and Douglas A. Macdonald, Black Holes, The Membrane Paradigm, Yale University, 1986.

[15] W. G. Unruh, "Experimental black hole evaporation," Phys. Rev. Lett. 46: 1351−1353 (1981), and "Sonic analog of black holes and the effect of high frequencies on black hole evaporation," Phys. Rev. D 51: 2827− 2838, (1995). As discussed in: http://www.scholarpedia.org/article/Hawking_radiation.

[16] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler, GRAVITATION, W.H. Freeman and Co., San Francisco, 1970, 1973. [MTW].

[17] "Derivation of Hawking Radiation," http://www.physicsforums.com/showthread.php?t=408124.

[18] 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].

[19] Wolfgang Rindler, Essential Relativity, 1969 [QC6 R477, p 171, also in MTW p 615].

[20] Benjamin Knorr, "Uniform Relativistic Acceleration, June 19, 2010, or "Uniform Acceleration," George Smoot, 2003. http://aether.lbl.gov/www/classes/p139/homework/eight.pdf http://www.physik.uni-leipzig.de/∼ schiller/ed10/Uniform%20relativistic%20acceleration.pdf
[21] The Future of Theoretical Physics and Cosmology, Celebrating Stephen Hawkings 60th Birthday, ed., G. W. Gibbons, et. al., Cambridge University Press, 2003 (879 pgs)"
[22] Dik Bouwmeester, et. al., "Experimental quantum teleportation," Phil. Trans. R. Soc. Lond. A (1998) 356, 1733-1737.
[23] Manfred Requardt , "On the Reality of the Unruh Temperature," [gr-qc] 7 Nov 2013, http://arxiv.org/pdf/1311.1608.pdf.
[24] Wikipedia, "Hawking Radiation," http://en.wikipedia.org/wiki/Hawking_radiation .
[25] Samir D Mathur, David Turton, "Comments on black holes 1: The possibility of complementarity," ArXiv 1208.2005 [hep-th] 11 Nov 2013. (Fuzzball complementarity). The fuzzball idea goes back to 2004.
[26] Samir D. Mathur, "What Exactly is the Information Paradox?" arXiv:0803.2030 13 Mar 2008.
[27] Fuzzball (string theory) http://en.wikipedia.org/wiki/Fuzzball_(string_theory).
[28] Ronald J. Adler, "Six easy roads to the Planck scale," arXiv:1001.1205 [gr-qc] 8 Jan 2010 (also appears in the American Journal of Physics).
[29] A. Zee, Quantum Field Theory in a Nutshell, Princeton, 2003. (e.g., p 265).
[30] Supermassive Black Holes, Wikipedia, http://en.wikipedia.org/wiki/Supermassive_black_hole

6. Notes:

Dark Energy Density:

The critical density of the universe is roughly
$\rho_{crit} = 3{H_o}^2/8\pi G \sim 10^{-26}\ kg/m^3 = 10^{-29}\ g/cm^3$.
Length, mass, and time can be converted into a common unit of electron volts, eV, using $1eV^{-1} = 1.97 \times 10^{-7} m$, $1eV\ mass = 1.78 \times 10^{-36}\ kg$, and time $eV^{-1} = 6.58 \times 10^{-16}\ sec$. This uses just the two basic constants: $c = 3 \times 10^8\ m/s$, $\hbar = 6.58 \times 10^{-16} eVs$. $1eV = 1.602 \times 10^{-19} J$.
Then $10^{-29}\ g/cc \sim 4.3 \times 10^{-11}\ eV^4$. Dark energy is about $\Omega_\Lambda \sim 0.7$, so we get:
$\rho_\Lambda \sim 3 \times 10^{-11}\ eV^4 \sim (2.4 \times 10^{-3}\ eV)^4$. It is a theoretical goal to explain this value. Roughly speaking, it is equivalent to about 4 hydrogen atoms per cubic meter of space.

Flamm Paraboloid:

The Flamm Paraboloid (1916) is a nice way to visualize the full Schwarzschild space by use of an embedding diagram facilitated by adding another dimension called the 'lifting-dimension, $z = z(r)$'. The Flamm Paraboloid is a curve revolved about the z-axis and is an outfacing parabola above the Schwarzschild radius with a spherical cap below $r = R$ [16]. It smoothly traverses the apparent horizon 'singularity,' and also allows one to picture the cause of perihelion shift as a small 'wedge deficit' [19]. We match the radial metric, g_{rr} with this new lifting metric for $ds^2 = [1 + (dz/dr)^2]dr^2 = [1 - 2m/r]^{-1}dr^2$ so that $(dz/dr) = (1 - 2m/r)^{-1} - 1$, or:

$$z(r) = \int_o^r \frac{dr}{\sqrt{r/2m - 1}} = \sqrt{8m(r - 2m)} + cnst \tag{15}$$

where $m = 2MG/c^2$.
This is also a motivation for a form of 'Rindler coordinates' where $\rho = 2\sqrt{2MG(r-2MG)} \sim z(r)$ with an altitude of zero at the horizon.

Coordinates for constant acceleration, g, in the z direction are best calculated using 4-vector algebra with 4-velocity $u = (\gamma c, \gamma\vec{v})$ (e.g., MTW, p 166). Results are: $t = (sinh(g\tau)/g,\ z = (cosh(g\tau) - 1)/g,\ \beta(\tau) = dz/dt = tanh(g\tau)$ [13]. The equation $(z-z_o)^2 - t^2 = 1/g^2$ plots as a hyperbola. And $dt^2 - dz^2 = [cosh^2(g\tau) - sinh^2(g\tau) = 1]d\tau^2$. Rindler coordinates are a little different from this. They use $T = z\,sinh(gt)$, $Z = z\,cosh(gt)$ so that $z = \sqrt{Z^2 - T^2} > 0$ and $ds^2 = -dT^2 + dZ^2$.
A better discussion of accelerated motion is found in web sources [20].

Black Hole Surface Temperature:
Wikipedia [24] (without saying so) implies that the temperature on the surface of a black hole is not the Planck temperature. Their sparse derivation of Hawking temperature (seen at 'infinity') is based on the Rindler lift coordinate above, $\rho = 2u = z(r)$, and the near-horizon observer sees a local inverse temperature:
(16)

$$\beta(u) = 2\pi\rho = 4\pi u = 4\pi\sqrt{2m(r-2m)} = 1/kT;\ \ so\ \ \beta(r') = 4\pi\sqrt{2m(r-2m)}\frac{\sqrt{1-2m/r'}}{\sqrt{1-2m/r}}.$$

transforming temperatures using the $\sqrt{g_{oo}}$ time component of the metric (as one should). Taking $r' \to \infty$ gives $g_{oo} = 1$, and we get, $\beta(\infty) = 4\pi\sqrt{mr}$. But r close to the horizon is $\sim 2m$, so we get $1/kT = \beta = 8\pi m$, or $T_H = \hbar c^3/8\pi MGk$. [14].
Let T_ℓ be the desired temperature on the stretched horizon.
Then $T_{r'}/T_r = T_\infty/T_\ell \sim \sqrt{M_p/M} \sim \sqrt{g_{oo}(\ell)}$.
Now, a one degree K black hole temperature corresponds to a mass near $1.2 \times 10^{23}\,kg$. So, $T_\ell \sim \sqrt{10^{23}/10^{-8}} \sim 10^{15}$!! But, the Planck temperature is near $10^{32}\,K$. This differs from Susskind's (ArXiv [9]) claim of near Planck temperatures on the stretched horizon.

$$T_\ell \sim T_{surface} \sim T_H\sqrt{\frac{2M}{M_p}} = \frac{T_p}{4\pi c}\sqrt{\frac{kT_p}{2M}} \tag{17}$$

For example, the smallest black hole at $M \sim 4M_\odot = 4 \times 1.99 \times 10^{30}\,kg$ with a Hawking temperature of $T = 16$ nano-Kelvins should have $T_\ell = 4.3 \times 10^{11}\,K \ll T_p$. The 20 billion sun BH should have horizon temperature of six-million kelvins.

One article claims to prove that the Unruh temperature is "real" [23]. It has key statements "that an accelerated detector will see a different kind of vacuum fluctuation pattern

[13]With c's included, the t eqn should have c/g, and the z eqn should use c^2/g. And recall that mass $m = MG/c^2$.
[14]This approach is also found in Antony Zee's QFT textbook [29]

compared to an inertial observer." And, "The crucial property that makes the relativistic regime so different compared to non-relativistic thermodynamics in an exterior gravitational field is that pure heat now has weight!"

RECENT RESULTS IN ASTROPHYSICS

DAVE

Abstract. The following is a collection of new key results from astrophysical observations. It includes the latest set of key cosmological parameters from CMB measurements, results from the Alpha Magnetic Spectrometer (AMS) on the Space Station, new determinations of the extragalactic background light (EBL), new data on galaxies, consideration of 'Warm Dark Matter' (WDM), and gamma-ray observed 'Fermi Bubbles.'

1. Cosmological Parameters from CMB

Cosmic microwave background radiation (CMB) fills the universe with the residue from the Big Bang, and space probes have been carefully measuring the the spatial anisotropies of its temperature. The previous results from the Wilkinson space probe (WMAP-2006) were: Hubble constant $H_o = 70.5 \pm 1.3\, km \cdot sMpc^{-1}$, $\Omega_{baryons} = 0.0456 \pm 0.0015$, $\Omega_{CDM} = 0.228 \pm 0.013,\ and\ t_0 = 13.72 \pm 0.12\, Gyr$ for the present age of the universe from the beginning of the Big Bang [4]. These values were recently updated from new additional measurements and analysis. The final high resolution CMB angular distributions posted by the WMAP collaboration all are still consistent with the ΛCDM model (black curve of Figure 1). Wiggles beyond the third peak were beginning to become clear, and new Planck data beyond WMAP makes them even clearer.

The European Space Agency's Planck satellite was launched on 14 May 2009 as the third in the series of CMB studies [5]. A group of 29 status reports came out in March, 2013 including an update on new constraints on inflation. The scalar spectral index is now $n = 0.9603 \pm 0.0073$ without any running. This rules out exact scale invariance at over 5σ; and some of the hundreds of inflation models are now ruled out. "Unless a quartic term is allowed in the potential, we find result consistent with second-order slow-roll predictions." Other results include slight changes in cosmic parameters such as $\Omega_\Lambda = 0.693$, $H_o = 68$, $Age = 13.8\, Gyr$. Baryon density has increased to 4.6% and dark matter to 26.8%. New detailed polarization data has not yet been published.

In addition, from the report on cosmological parameters [6], "The results from Planck are consistent with the results of standard big bang nucleosynthesis. In fact, combining the CMB data with the most recent results on the deuterium abundance, leads to the constraint $N_{eff} = 3.02 \pm 0.27$ " for the number of neutrino species. A summary of

---

Date: 5 September, 2012, paper updated to 24 March 2014.

email: davepeterson137@gmail.com. Boulder, Colorado.

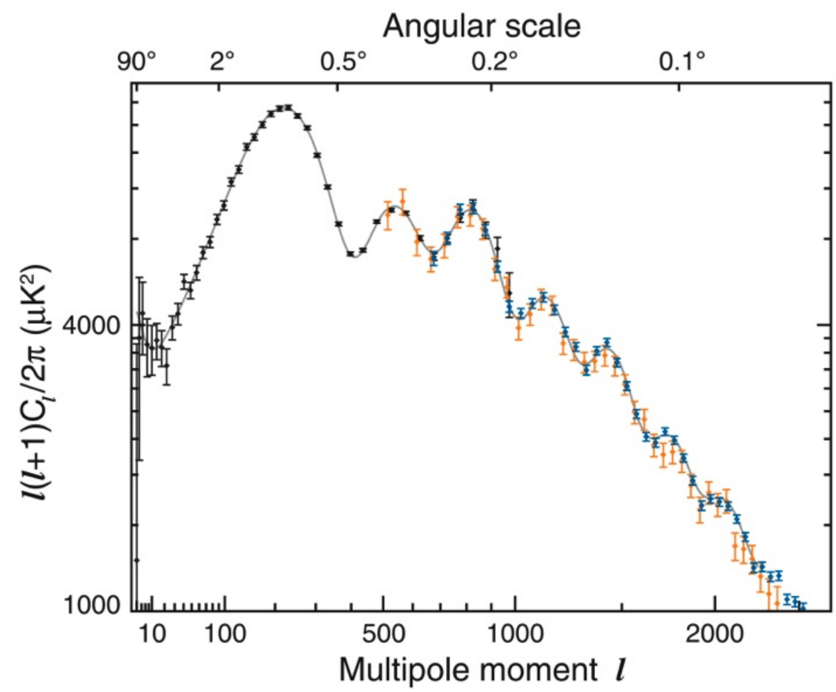


FIGURE 1. High resolution of Cosmic Blackbody angular peaks (WMAP 9-year update).

the angular correlation statistical analysis is that "at high multipoles, the base ΛCDM cosmology provides an excellent fit to the spectra, but the parameters derived from the CMB apparently conflict with some types of astrophysical measurement. From an analysis of an extensive grid of models, we find no strong evidence to favour any extension to the base ΛCDM cosmology, either from the CMB temperature power spectrum alone, or in combination with the Planck lensing power spectrum and other astrophysical data sets."

The Planck space telescope was turned off on 23 October, 2013, after gathering data for four and a half years (from 2009). It should now be out of helium, it is no longer transmitting, and it will be parked in a distant orbit around the sun (permanent hibernation).

2. Inflation

In March, 2014, after a three-year analysis, the South Pole telescope group BICEP2 [10] announced its discovery of what is called B-Mode Polarization (at more than 5 sigma confidence above expected background contamination). This was a difficult experiment because the B-mode signal is only about one-percent as strong as the previously analyzed weak background temperature fluctuations .The "Background Imaging of Cosmic Extragalactic Polarization study number two (BICEP2) revealed pinwheel-like swirls in the polarization

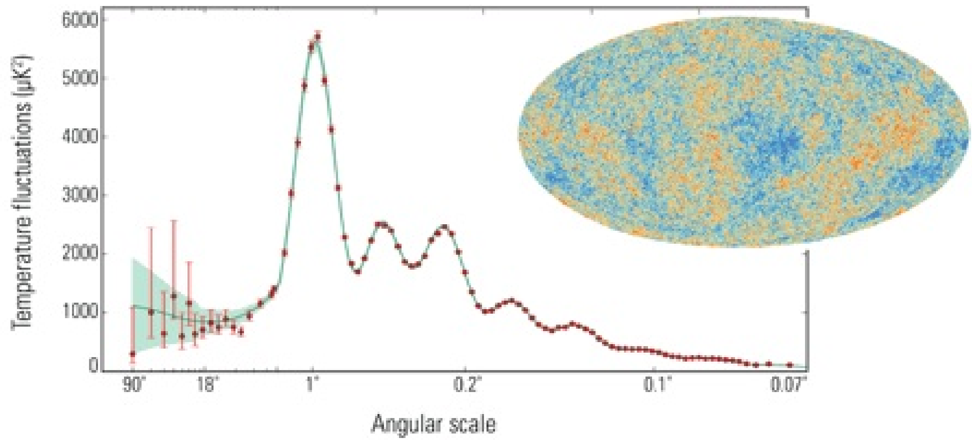


FIGURE 2. High resolution of Cosmic Blackbody angular peaks (Planck Satelite). 6-parameter curve fit.

of the cosmic microwave background (CMB). The parameter expressing this curling polarization is called "r, the ratio of power in tensor to scalar density perturbations of the CMB. Its value is $r = 0.20$ and represents a much higher signal than anticipated. Over the next year, it should not be difficult for a variety of other science experiments to verify this value. This discovery is considered to be major because it may be strong evidence for the early inflationary expansion of the universe. That GUT scale blow-up has powerful gravitational waves which produce the circular polarizations. A previous belief in "Higgs Inflation is now dead because it suggests $r = 0$ instead. And the ekpyrotic model is also now untenable. What is left is a standard inflation from some new GUT field near 10^{+16} GeV as in modeling by Andre Linde. Still, further science is needed because there is a discrepancy between these Antarctic results and the present analysis from PLANCK which suggests a highest value of $r < 0.11$.

3. AMS

The abstract from a new 2013 paper says, "A precision measurement by the Alpha Magnetic Spectrometer on the International Space Station of the positron fraction in primary cosmic rays in the energy range from 0.5 to 350 GeV based on positron and electron events is presented. The very accurate data show that the positron fraction is steadily increasing from 10 to 250 GeV, but, from 20 to 250 GeV, the slope decreases by an order of magnitude. The positron fraction spectrum shows no ne structure, and the positron to electron ratio shows no observable anisotropy. Together, these features show the existence of new physical phenomena [7]." Modeling fits suggest that "a significant portion of the high-energy electrons and positrons originate from a common source." these features show evidence of a new physics phenomena. The exact shape of the spectrum, extended to higher

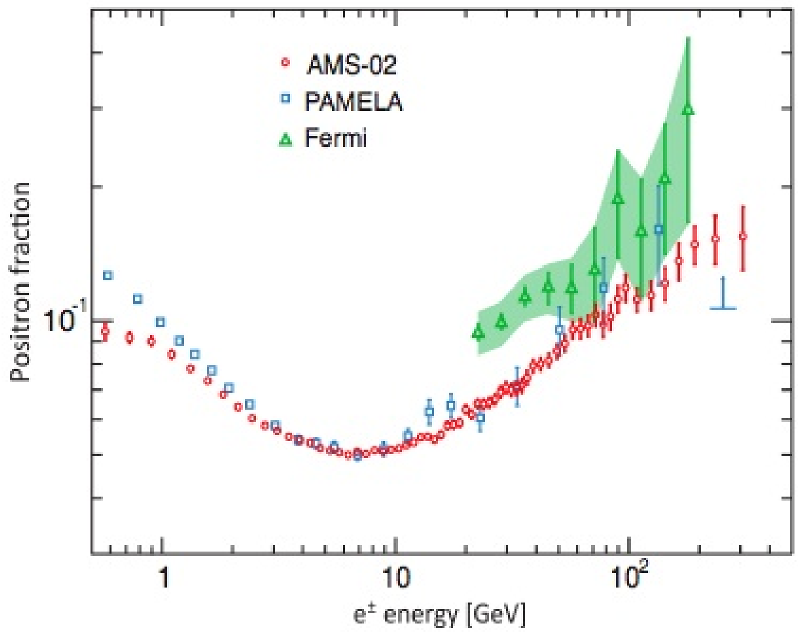


FIGURE 3. Plot of positron fraction $e^+/(e^- + e^+)$ against energy (GeV) for new AMS data versus previous data from PAMELA and Fermi.

energies, will ultimately determine whether this spectrum originates from the collision of dark matter particles or from pulsars in the galaxy. The high level of accuracy of this data shows that AMS will soon resolve this issue.
Despite the new wonderful resolution of the AMS, the central ideas were already known from the previous PAMELA and FERMI satellite data. We already knew about the increase and that some new source of positrons must exist above 10 GeV energy.

However, there is no current reliable theory for the propagation of background cosmic rays; and it is necessary to eliminate this background before making a claim of positron excess due to dark matter [16]. Along with positrons, a background of antiprotons and cosmic ray nuclei can also be produced in secondary collisions with ambient gas in the galaxy. A recently calculated upper limit for the naturally occurring positron fraction lies above the currently measured positron fraction. Therefore, no new excess can be stated as being due to either pulsars or dark matter. The measured antiproton flux is in agreement with the calculations.

In addition, it is suggested that if WIMP dark matter actually exists, its mass must be above 26 GeV (2σ confidence) [17]. This claim comes from analysis over the full CMB data base. It has also been reported that no findings of direct detection (LUX) have been seen. So, AMS needs to come up with additional higher energy positron data. And the general

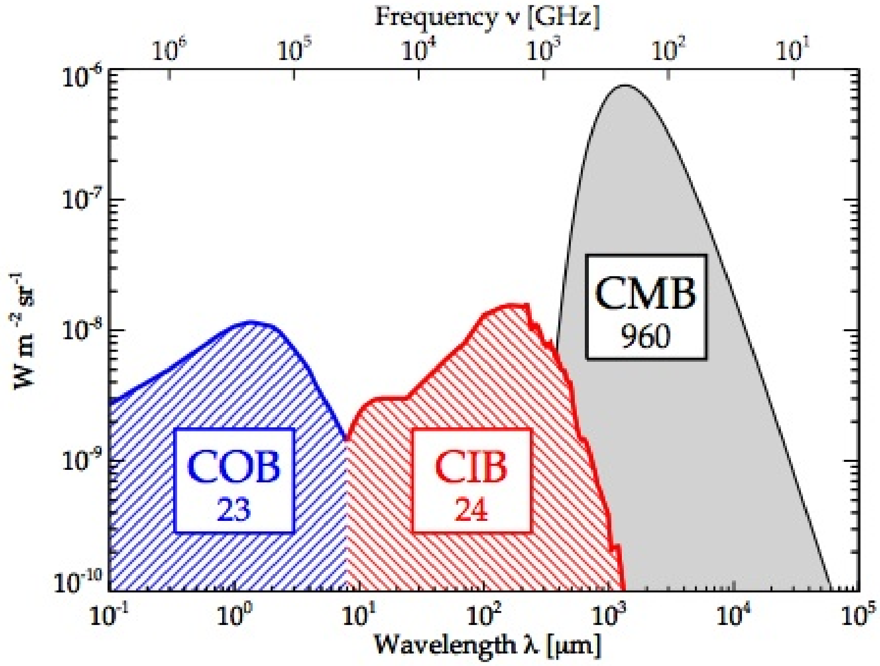


FIGURE 4. Spectral Energy Distribution for background optical, infrared and microwave photons. Numbers refer to the integrated power in each peak in nW/m^2sr.

understanding cosmic ray physics needs to be improved.

4. How much Light from Galaxies since the Big Bang?

Researchers can now see extragalactic background light (EBL) that fills the universe by measuring the attenuation of high-energy-gamma rays from distant blazars [8]. This EBL is made up of all the ultraviolet, optical and infrared photons ever emitted by all galaxies in the universe summing over the red-shifts for all cosmic star formation history. Blazars are compact quasars from supermassive black holes in the nuclei of active galaxies (BL-Lac AGNs) which also happen to have relativistic plasma jets pointing in our direction. These jets contain very high energy gamma rays (VHE $\sim 30\,GeV - 30\,TeV$) which can interact with the EBL to produce electron-positron pairs detectable directly using Cerenkov telescopes or indirectly via attenuation. That can happen, for example, when a 2.6 TeV gamma ray hits a background 0.1 eV EBL photon. It is now clear that distant blazar high energy photons are attenuated more than those from nearby blazars. Estimates of EBL have been determined previously by direct observations but were suspect due to significant contributions from the light from our own Milky Way galaxy. New deductions from this photon-photon pair production attenuation give results that agree with previous direct observations of EBL photons.

Background power can be broken into 'optical' versus 'infrared' for photons with wavelength less than 8 microns and those greater than 8 microns up to a millimeter. These are labeled as COB (cosmic UV and optical starlight background containing a total of 24 nW/m^2sr) and CIB (cosmic infrared dust-reprocessed-starlight background with 23 nW/m^2sr) [9]. These approximately equal totals are only 5% of the total background with CMB microwave photons having a much larger power of 960 nW/m^2sr. The peak power for COB is at 1.3 μm versus 150 μm for CIB. This means that there are approximately 115 IR photons for each visible photon. (ref [9] using direct integration of EBL light). Along with these determinations are estimates of the gamma ray opacity of the universe (CGRH = cosmic gamma ray horizon) or optical depth of the EBL versus synchrotron models for VHE flux without attenuation (see Dominguez [8]). This can be characterized by an e-folding distance for attenuation. Plots can be made for the evolution of EBL back to a redshift of $z = 0.5$ or 5 billion years ago. A few examples from the plotted curve are: at nearby $z \sim 0.01$, 10 TeV γ rays can barely survive. At redshift of $z \sim 0.1$ the CGRH is near 1 TeV, and at $z \sim 0.5$, CGRH is near 200 MeV. The universe is opaque against higher energy gamma rays than these.

5. Galaxy

The rotational star-flow of the Milky Way galaxy has recently been determined near the Sun's location using the "Apache Point Observatory Galactic Evolution Experiment" (APOGEE) data. The result is an approximately flat circular velocity of $V_c(R_o) = 218 \pm 6\ km/s$ at a radius near $R_o \simeq 8.5\ kpc \pm 0.5$. The sun itself moves with respect to the flow with a radial velocity near $-10 \pm 1\ km/s$ and local circular velocity near $+26\ km/s$[1].

There are now two central milky way stars whose full orbits have been measured about our mega-black hole. These are star SO-2 with a 16.5 year orbit and now SO-102 with an 11.5 year orbit about $Sgr\ A^*$ whose mass is 4 million solar masses [2]. These objects were observed by the Keck telescope at Mauna Kea in Hawaii from 1995 to 2012. Other stars also seem to be orbiting but only partial orbits have been recorded so far. General relativistic effects should be verified in the future (red-shift at periapsis, orbital overshoot per period).

6. Planes of Galaxies

The small satellite galaxies surrounding Andromeda (M31) and the Milky Way do not lie in a spherical distribution as often suggested by ΛCDM. Instead, 27 dwarf galaxies about Andromeda lie in a thin disk, and 24 small galaxies about the Milky Way also lie in a plane called "The Great Pancake" [18]. The vast thin plane of dwarf galaxies (VTPD) orbiting M31 became clear in 2013 (Ibata), and the thin plane of satellites roughly perpendicular to

the Milky Way disk was reported in 2005. In addition to this, about 14 of the local large galaxies also lie in a plane (a "Local Sheet Council of Giants 34 Mly across and only 1.5 Mly thick) [19]. This pertains to all bright galaxies within about 20 Mly away including us. Twelve of these are spiral galaxies.
Some simulations suggest that the satellite planes may be due to infall along spines of filaments of the cosmic web. Having co-rotating satellites in a thin plane for both M31 and MW does suggest similar formation history. A "Millennium-II simulation might show consistency with ΛCDM cosmology.

7. Warm Dark Matter

Despite previous successes and even incorporation into the name "Standard Model of Cosmology," cold dark matter (CDM) cosmology is now seen to badly fail to properly model the real world at the scale level of galaxies (1-100 kpc). CDM (e.g., WIMPs, neutralinos, $\sim$ 10 GeV particles) yields too much local structure. Of course, hot dark matter (e.g., neutrinos) is too light to express cosmology or give any local structure at all. In-between is a new scenario of warm dark matter (WDM, e.g., right handed 'sterile neutrinos, gravitinos, axinos , near 1 keV energy $-$ sometimes also called 'keVins (keV level inert fermions [12]).

For large scales ($\geq$ 100 kilo-parsecs), CDM and WDM can both agree with cosmology, but it is WDM that can best agree with galaxies. [However, there are many different theories for CDM and WDM so that CDM cannot yet be clearly excluded]. WDM has now 'attracted considerable attention.

The problems with CDM include: 1) WIMPs have never been seen (at LHC nor underground experiments), 2) CDM predicts an abundance of galactic satellites (the famous "missing satellite problem"), 3) the 'CUSP problem' (narrow density profiles rather than broad cores for galaxies), 4) 'too big to fail' (massive failure problem or MFP [prediction of massive subhaloes of the milky way of high concentration and circular velocity that cannot host bright satellites and are not observed]). There is prediction of galactic bulges (but pure disk galaxies do exist without bulges). WDM subhalos have the right concentration to host the bright Milky Way satellites. Observed substructures suggest a $\sim$ 2 keV dark particle such as those predicted for sterile neutrinos (Pontecorvo, 1968), and there are now many present experiments to search for sterile neutrinos. With CDM, a halo of dwarf galaxies cannot form the way they actually do. There are also problems with large scale velocity flows, voids, fainter 1a supernovas at high z, poor Tully-Fisher relation, poor galaxy Bar stability, and too short 'free streaming lengths'. Present N-particle simulations are working better with WDM than with CDM.

Like strings, ΛCDM now has its own internal inertia. But,"Putting all together, evidence that ΛCDM and its proposed baryon cures do not work at small scales is staggering. Increasing and impressive evidence favour a fermionic DM particle mass of about 2 keV which naturally produce galaxy observations, cored density proles and their sizes. Quantum WDM effects are important, particularly for dwarf galaxies. Overall, ΛWDM and keV scale DM particles deserve dedicated astronomical and laboratory experimental searches, theoretical work and numerical simulations" [11].

8. Fermi Bubbles

Feedback mechanisms exist that can put the brakes on galaxy formation. One of the most important of these results from the growth of super-massive-black-holes (SMBHs). Liberated energy from accreting mass can accelerate the polar dispersion of gas to very high speeds away from the plane of the galaxy [13]. Many Active Galactic Nuclei (AGNs) have been studied largely from their radio emissions. Initially, gas is pushed away violently from the AGN but then gradually at lower speeds until the gas is gently blowing bubbles over giant lobes. It is also known that starbursts can drive enormous gas outflows.
The center of our own galaxy, the Milky Way, has an SMBH called Sgr A^* with a mass of over four million suns. It is presently inactive; but millions of years ago, it could have had a period of activity. In addition to the black hole, the central volume of the galaxy is very dense with stars and gas. Within just one light year of center there are over 100,000 stars − much denser than our local region. A thick layer of dust and gas and background radiation has prevented us from seeing into the center until fairly recently. WMAP microwave observations of the CMB saw excess microwave haze near galactic center, and this was confirmed with more recent Planck data. Radio telescopes have detected two giant bi-polar supersonic outflows of charged particles via linearly polarized radio lobes. Hints of giant bubbles and their edges were also seen in X-ray observations from the German Roentgen Satellite (ROSAT).

Most spectacularly, in May of 2010, two giant gamma ray bubbles were seen extending 25,000 light years north and south of the center of our Milky Way galaxy (see Figure 5). The first firm observation came from NASA's Fermi Large Area Gamma-ray Space Telescope (LAT, and a previous name was GLAST). Observations include photon energies from 1-10 GeV. The huge structure spans more than half of the visible sky, from the constellation Virgo to the constellation Grus, and it may be millions of years old. One initial belief was that these lobes derived from a period of Quasar activity from Sgr A^*. The observed energy spectrum seems to be due to gamma-ray photons generated by inverse Compton scattering processes from a population of relativistically moving electrons. It is believed that observations over all energies (radio, microwave, visible, x-ray, gamma-rays) are explained by similar phenomena.

The observed radio lobes [14] correspond to the gamma-ray Fermi bubbles and possess a strong magnetic field up to 15 microgauss. The analysis of the radio telescope report says that the lobes are driven by star-formation activity rather than black-hole driven outflow

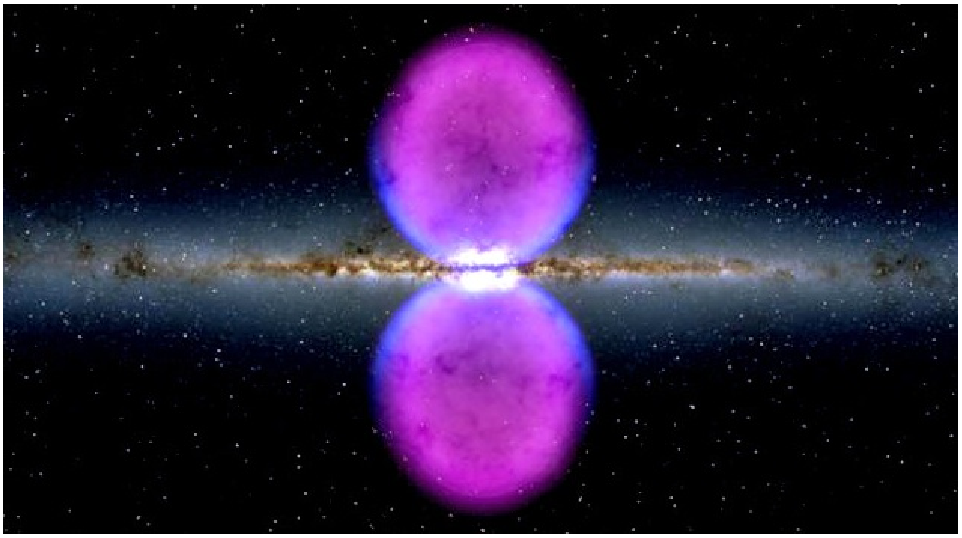


FIGURE 5. Appearance of gamma-ray lobes above and below our Milky Way galaxy

from galactic center. This intense star formation suggests more than a ten-million year history of sequential supernovae explosions. There is still much uncertainty about the cause of the Fermi lobes. Possibilities include the collision between two black holes about ten million years ago. One author thinks that the gamma ray spectrum could come from dark matter annihilation of 62 GeV particles. There are claims of jets, but this seems unlikely due to the high symmetry of the lobes. One statement was that the jets were tilted at an angle of 15 degrees from perpendicular. The accreted mass required to power the lobes must be at least 2000 suns [15].

References

[1] Jo Bovy, et. al., "The Milky Way's Circular Velocity Curve Between 4 and 14 KPC from APOGEE Data," ArXiv:1209.0759v1 [astro-ph.GA] 4 Sep 2012

[2] L. Meyer, et. al., Science 5 Oct 2012 338 no 6103 p 84-87, and see "Star seen whizzing around supermassive black hole," Physics World, 10/4/12.

[3] WMAP Collaboration post, "How many neutrinos in the Sky?", http://resonaances.blogspot.com/ 18 Jan 2013.

[4] E. Komatsu, et. al. "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP1) Observations: Cosmological Interpretation." For AJP Supplement 8/13/06 [Preprint at: http://lambda.gsfc.nasa.gov/product/map/dr3/pub_papers/ fiveyear/cosmology/wmap_5yr_cosmo.pdf]

[5] Planck Collaboration, "Planck 2013 results. XXII. Constraints on Inflation," arXiv:1303.5082 [astro.CO] 20 Mar 2013, also see "Overview" 1303.5062 listing all XXIX reports so far.

[6] Planck Collaboration, "Planck 2013 results. XVI. Cosmological Parameters," arXiv:1303.5076, 20 Mar 2013.

[7] M. Aguilar, et. al., "First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5350 GeV," Physical Review Letters, 5 April 2013, PRL 110, 141102 (2013) AMS Collaboration (including Samuel Ting).

[8] A. Dominguez, et. al., "Detection of the cosmic γ ray horizon from multiwavelength observations of blazars," ArXiv 1305.2162 and The Astrophysical Journal , 2013 (also see discussion at Physics World, 5/27/13). And see Science Daily May 24, 2013, "Detection of the Cosmic Ray Horizon: Measures all the light in the universe since the big bang." Previous publication at: ArXiv:1103.4534 astro-ph 15 Mar 2011.

[9] H. Dole, et. al., "The Cosmic Infrared Background Resolved by Spitzer," ArXiv:astro-ph/0603208 8 Mar 2006 for Astronomy & Astrophysics, 2008 (primary graph later used by ArXiv:1304.1330).

[10] BICEP2 http://bicepkeck.org/ BICEP Release package. ArXiv1403.3985 "BICEP2: Detection of B-mode Polarization at degree angular scales." http://arxiv.org/pdf/1403.4302.pdf "BICEP2: II: Experiment and Three-Year Data Set," Mar 17,2014, astro-ph. Also see Planck 1403.3919.

[11] "From Cold Dark Matter to Warm Dark Matter in the Standard Model of the Universe Theory and Observations," Paris Cosmology Colloquium, ArXiv 1203.1874 7 Jul 2013.

[12] H. J. de Vega and N. G. Sanchez, "Dark Matter in Galaxies: the Dark Matter Particle Mass is about 2keV," arXiv 1304.0759 2 Apr 2013.

[13] Mitchell Begelman and Martin Rees, Gravitys Fatal Attraction, Black Holes in the Universe, Second Ed., Cambridge Universe Press, 1995, 2010.

[14] Ettore Carretti, et. al., "Giant magnetized outflows from the center of the Milky Way," Nature 493, 66-69, 03 Jan 2013.

[15] K. Zubovas, et. al., "The Milky Ways Fermi Bubbles: Echoes of the Last Quasar Outburst" (for MNRAS) 28 Apr 2011. Also on ArXiv: 1104.5443

[16] Kfir Blum, Boaz Katz, and Eli Waxman, "AMS-02 Results Support the Secondary Origin of Cosmic Ray Positrons," Phys. Rev. Lett. 111, 211101 (2013) [5 pages]. Also see, "Does the positron excess really exist." Physics World article 11/27/13, http://physicsworld.com/cws/article/news/2013/nov/27/does-the-positron-excess-really-exist

[17] Mathew S. Madhavacheril, Neelima Sehgal, Tracy R. Slatyer, "Current Dark Matter Annihilation Constraints from CMB and Low-Redshift Data, http://arxiv.org/pdf/1310.3815v1.pdf

[18] Noam Libeskind, "Dwarf Galaxies and the Dark Web," Scientific American, March 2014, p 47-51. Along with: Noam I Libeskind, et. al., "The Distribution of Satellite Galaxies: The Great Pancake," http://arxiv.org/pdf/astro-ph/0503400.pdf, 2005, for Mon. Not. R. Astron. Also see: Rodrigo A. Ibata, et al., "A Thousand shadows of Andromeda: Rotating Planes of satellites in the Millennium-II cosmological simulation," ArXiv: 1403.2389 astro-ph 10 Mar 2014.

[19] "Milky Way amidst a Council of Giants." Science Daily 3/12/14 for publication in MNRAS, 2014, M. L. McCall.

THE LAST DECADE IN EXPERIMENTAL PARTICLE PHYSICS

DP

Abstract. The last decade of particle physics has been largely a time of slow progress, development of basic themes, and some new discoveries. The latest important discovery is the announcement in July, 2012 that the existence of a new particle resembling the Higgs boson at 126 GeV is firm from both ATLAS and CMS at the CERN Large Hadron Collider (LHC). Otherwise, recent history has not been quite as exciting as say the **"November Revolution"** of 1974 when charmonium was discovered with subsequent realization over the next few years that the quark model was valid and the standard model took shape (e.g., 1974 J/ψ meson, 1975 tau lepton , 1977 Upsilon, 1979 gluon jet, 1983 W and Z weak bosons). However, we are now in the midst of another very interesting **'Neutrino Revolution'** from the observation of flavor changing neutrinos. A primary goal has been the search for new physics beyond the standard model, a search that is still in progress mainly at the CERN LHC. This 27 km-circumference proton collider was switched on in September, 2008 after 25 years of planning and construction. So far, most new reports say something like, "No clear excess above the Standard Model expectations is observed." LHC is also performing RHIC type experiments to clarify the nature of the high energy quark-gluon plasma, and some new discoveries have been made in that arena.

1. The Higgs Boson

The LHC was designed to explore the breaking of electroweak symmetry and search for physics beyond the Standard Model (SM). But, by the end of 2011, there had not yet been any firm experimental evidence that the Higgs particle existed. Many possible decay modes had been investigated, $H \to \gamma\gamma, \tau\bar{\tau}$, b$\bar{b}$, W^+W^-, or Z$\bar{Z}$. What was called preliminary data from most major tests and channels indicated that a Higgs boson would be shown to exist near a mass of 125 GeV with confidence level above 3 standard deviations. At least the higher candidate range of $m_H \simeq$ 129-525 GeV had been excluded. During 2012, the LHC was able to run at a slightly higher 4 TeV/beam with about 3 times higher luminosity, and decent Higgs statistics was finally obtained.

<u>July 4, 2012: CERN</u> CMS and ATLAS announced the official existence of a new particle with properties similar to that expected for the Higgs particle and having a mass of 126 GeV with essentially 5-sigma confidence. This fulfills a major goal of the LHC. There are still details to be worked out about all the higgs decay modes and how it fits into expected physics. Figure 1 shows the enhancement bump near 125 GeV for the strong-decay diphoton channel for CMS data.

Date: November 15, 2011. Paper updated to Feb 12, 2014.

email: davepeterson137@gmail.com. For BPL Cosmology group.

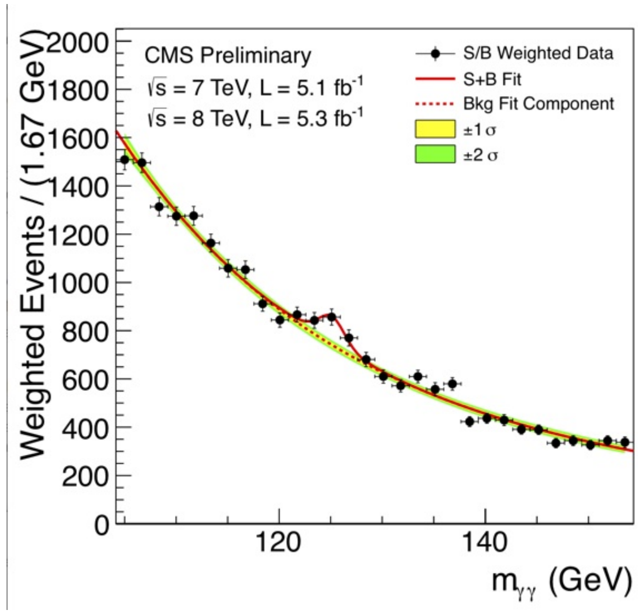


FIGURE 1. "Higgs particle Bump" Experimental results from 2011/2012 LHC proton-proton collision data in the two-gammas (di-photon) channel.

The existence of an $m_H \simeq 126$ GeV definite particle boson resembling the Higgs was the last outstanding piece of the standard model (SM). The Higgs field permeates all space, and its interactions with elementary particles gives them mass by providing a condensate that can break gauge invariance. The Higgs mechanism is a type of weak-force superconductivity of the vacuum (similar to that which gives effective mass to the photon in electrical superconductors). One over-simplified picture of its action is to provide a pool of 'molasses' that can stick to particles traveling through it. Different particles interact with different strengths, but the top quark interacts most strongly (and has a higher mass, $m_t =$ 177 GeV, than the Higgs). One curiosity is that the Higgs mass is the geometric mean of the top quark mass and the Z boson mass to better than one percent, ${m_H}^2 \simeq m_t m_Z$.

The Higgs boson is the excitation quantum of the Higgs field, and its identification is seen from decays into two photons and into two Z's with predictions agreeing with the standard model for Z's but perhaps somewhat higher than expected for di-photons. Like the Ws, the Higgs decays quickly, and hence its detection 'bump has a wide energy width. Higgs particle production derives mainly from gluon-gluon fusion, $gg \rightarrow H$ [1], from loop diagrams mediated by quarks but mainly using the top quark. The decay $H \rightarrow \gamma\gamma$ isn't that common and also depends on loop diagrams mainly from W's. Decays into the W's channel and also the tau's and b's weakly seems to be below expectations so far. More

[1] "The fact that this Higgs-like particle is produced via gluon fusion provides indirect evidence of its coupling to top-quarks. The coupling to other fermions is uncertain." [ArXiv:1212.0560]

data is needed for clarity, and higher energy is needed for self-interaction of the Higgs with itself. Many physicists are applying their favorite theories to the current meaning of the Higgs and its decay modes with some form of SUSY being the main hope [28]. The original intention of the Higgs field was to give mass to the weak bosons. It was then assumed that this mechanism would also give mass to the fermions as well. Since the Higgs doesn't seem to be decaying into tau leptons, it may not be giving mass to them either. Then maybe fermion mass comes from new SUSY Higgses instead. The di-photon Higgs bump appears atop a background produced by random photon pairs. ATLAS and CMS hint at roughly a $+2\sigma$ overproduction which may be interesting if the statistics continue. One possible alternative beyond the standard model is MSSM (minimal supersymmetric SM) which has two complex Higgs doublets − sometimes called H_d and H_u. This gives $2 \times 4 = 8$ fields with 3 being 'eaten' up for the weak bosons leaving 5 Higgs bosons called $h, A,\ H^+, H^-, H^o$. There is no current evidence for this proposal.

Bill Ford (CU) believes that with current statistics, all is within expectation. Details will continue to emerge from CERN over the next two years. The perhaps bigger news from CERN is the apparent absence of supersymmetry (SUSY − again, so far). Physicists are hoping to see the 'stop' or susy-top-quark-superpartner. If the higgs turns out to be too *normal*, then physicists are in a quandary about what might lie around the next bend. Only about one H particle is produced per 10^{10} pp collisions, but 10^{15} pp's have occurred so far. Note that current claims for the Higgs are mainly based on easy but relatively rare decay modes called the "diphoton" mode ($H \to \gamma\gamma$, 0.2%) and the ZZ *or* 4ℓ or 4-lepton mode ($\simeq 3\%$). The much bigger modes have not yet been reported: "bottom-antibottom" (57%) of H decays, W^+W^-(21%), gluon-gluon (9%), and $\tau\bar{\tau}$, (6%). [32]. The H doesn't have enough mass to make two Z's or 2 W's, so one must be virtual and short lived.

The Weinberg-Salaam ElectroWeak (EW) theory requires four scalar fields − three of which are used up in making the massive W's and Z, and one to give a "Higgs particle," H. A nice presentation of this is given on Matt Strassler's website [22]. The discovery of the Higgs favors actual physical use of EW scalar fields rather than new strong "technicolor forces." "This is why the discovery is important." [Weinberg, July, 2012].

2. Neutrino Oscillations

The **'Neutrino Revolution'** refers to the observation of flavor changing neutrinos after 1998. There are three different types or flavors of neutrinos (ν_e, ν_μ, ν_τ, and their antiparticles: $\bar{\nu_e}$, $\bar{\nu_\mu}$, $\bar{\nu_\tau}$). The existence of the electron neutrino was experimentally verified in 1956 from nuclear reactors, the muon neutrino was seen in 1962, and the tau neutrino in 2000. On earth, the main source of electron neutrinos is from nuclear reactions in the sun. Neutrinos are also produced from inside the earth and in the atmosphere from cosmic rays that also yield muon neutrinos. And then of course there are the greatly numerous cosmic background neutrinos which now have very low energy. These will be ignored here because they are unlikely to ever be detectable.

Detections of solar electron neutrinos occurred from 1970 to 1994 a mile below ground in an old South Dakota gold mine. Ray Davis (Nobel prize 2002 at age 88) and John Bahcall ran an experiment using 100,000 gallons of dry-cleaning fluid and counted rare individual argon atoms converted from chlorine by the neutrinos from the sun. The mysterious net result was that only a third of the expected count was observed − the "solar neutrino problem." The solution to the mystery is that electron neutrinos from the sun are converted into muon and tau neutrinos so that the net abundance at earth is about the same for all three types. The first proof of neutrino oscillation (transmutations and tiny neutrino masses) occurred in 2001 from an underground detection tank in Canada ("SNO") using 1000 tonnes of heavy water surrounded by 9600 photo-multiplier tubes. It could see all types of neutrinos, and their total flux finally agreed with solar theory. The first SNO (Sudbury Neutrino Observatory) experiments of June 2001 indicated that the solar neutrino problem was due to particle physics rather than solar astronomy. Solar electron neutrinos from the decay of 8-Boron were seen along with elastic neutrino scattering in heavy water which sees all types of neutrinos.

After 2001, a human controlled experiment ("K2K") verified the loss in flight of muon neutrinos at "Super-Kamiokande" Japan. This transmutation loss was verified in 2005 by a Fermilab-to-Minnesota ("MINOS") experiment. The K2K test used muon neutrinos created from the "KEK" synchrotron beamed through the earth over 250 km to a detector using 50,000 tons of water. A different later experiment called T2K (Tokai to Kamioka, Japan) in 2011 showed that some muon neutrinos can interconvert into electron neutrinos[1].

That the electron neutrino survival probability really does oscillate with distance traveled was clearly demonstrated by the experiment "KamLAND" in 2002 (see Fig. 2). Antineutrinos from nuclear power plants showed a sine wave probability with flight distance L (survival versus L/Energy). Neutrinos can be created in the atmosphere from cosmic ray collisions, and muon neutrinos from above are more plentiful than muon neutrinos coming from below. In traveling through the earth, many muon neutrinos seem to change into tau neutrinos − a loss first noted in Japan in 1998. The physics of neutrinos is a work in progress with many unanswered questions. There are many large experiments under construction or awaiting publication and much more to be discovered.

One of the most important mixing angles in the neutrino sector, θ_{13}, "has been shrouded in mystery for a long time." [7]. But then the experiments called T2K, Double Chooz, and MINOS hinted at a large $\sim 10^o$ non-zero PMNS matrix angle angle associated with loss of 'inverse beta decay' signals between near and far detectors of electron anti-neutrinos from nuclear reactors (using $\bar{\nu_e} + p \rightarrow e^+ + n$ reaction). Finally, in March, 2012, a precise measurement was achieved: "The Daya Bay Reactor Neutrino Experiment at Guangdong, China has measured a non-zero value for the neutrino mixing angle θ_{13} with a significance of 5.2 standard deviations [20] using antineutrinos from six reactors with 55 days of data. The latest result is $\theta_{13} \simeq 8.8 \pm 0.8^o$ ($\pm 1\sigma$ range). This very interesting non-preservation is much stronger than the CKM quarks case ($\theta_{13} \simeq 0.2^o$). The neutrino matrix 'describes a fundamental mismatch between the weak-interaction (flavor) and mass eigenstates of six leptons.' A previous analogy to this phenomenon was the oscillation of K-mesons due to a mismatch between the mass eigenstates and flavor eigenstates of kaons. The mass bases were called the "short" and "long" kaon state, and the flavor states were the neutral kaon K^o and its antiparticle. The flavor states are actually the ones that get measured and are plus and minus superpositions of the mass states with phases which evolve over time. So, one can start with a $|K^o\rangle$ and then end up with a $|\bar{K}^o\rangle$ [quark view: $(d\bar{s}) \to (s\bar{d})$] .

In the CKM case, there was a relationship between angles and the strong hierarchies of quark masses (e.g., Cabibbo $\sin\theta_c \simeq \sqrt{m_d/m_s} = \sqrt{4.79MeV/92.4MeV}$) [$\theta_c = \theta_{12}^{CKM} \sim 13.1^o$]. A similar analogy for leptons might say $\sin\theta_{23} \simeq \sqrt{m_\mu/m_\tau} + \sqrt{m_2/m_3} \simeq 0.65-$ actual result $\theta_{23} \simeq 45^o$. Also note that $\theta_{13}^{PMNS} \sim \theta_c/\sqrt{2}$, as suggested by several GUT models beyond the standard model [21]. It could also be that $\theta_{12}^{PMNS} + \theta_c = 45^o$.

At present, neutrinos are only left handed with no evidence for right handed spins ('sterile' neutrinos, ν_s) [4]. That is, if your left hand fingers curl in the direction of spin, then your thumb points in the direction of motion near the speed of light. Anti-neutrinos are only right handed, and the ability to convert directly from muon neutrinos to electron neutrinos is not yet established. It is established that only three light neutrinos can exist − but possible heavy neutrinos are not eliminated. It is not known whether massive neutrinos are also their own antineutrinos (Majorana neutrinos) or whether CP (charge-parity) violation occurs. It does now seem likely that Majorana particles will soon turn up in solid state physics.

"The quantities which are <u>not yet determined</u> are: the pattern of mass hierarchy (in other words the sign of Δm_{31}^2, the CP phase δ, and the 'octant' of θ_{23} (in other words the sign of $\theta_{23} = \pi/4$). [2]

<u>Some Other Dates:</u>
<u>Aug 16, 2007</u>: First real time detection of Be7 $\to$ Li7 solar neutrinos by Borexino! (Gran Sasso underground laboratory in Italy).
Aug 21, 2008: Measurement of the solar 8-B neutrino flux with 246 live days of Borexino and observation of the 'MSW' [3] vacuum-matter transition.
Mar 01, 2010: Observation of Geo-Neutrinos (from deep inside the earth). Borexino pseudocumene (trimethylbenzene) scintillator which has a much higher light output than mineral oil based liquid scintillators.
2008-2012: finally verified 'sun-like' fusion of proton+proton + electron 'pep' reactions with 1.44 MeV ν's at Borexino − in agreement with theory.

[2] ['Octant' rather than 'quadrant' because of three dimensions 123, *ArXiv:1211.7175, 11/30/12*].
[3] Mikheyev-Smirnov, Wolfenstein 1978 matter electrons effect

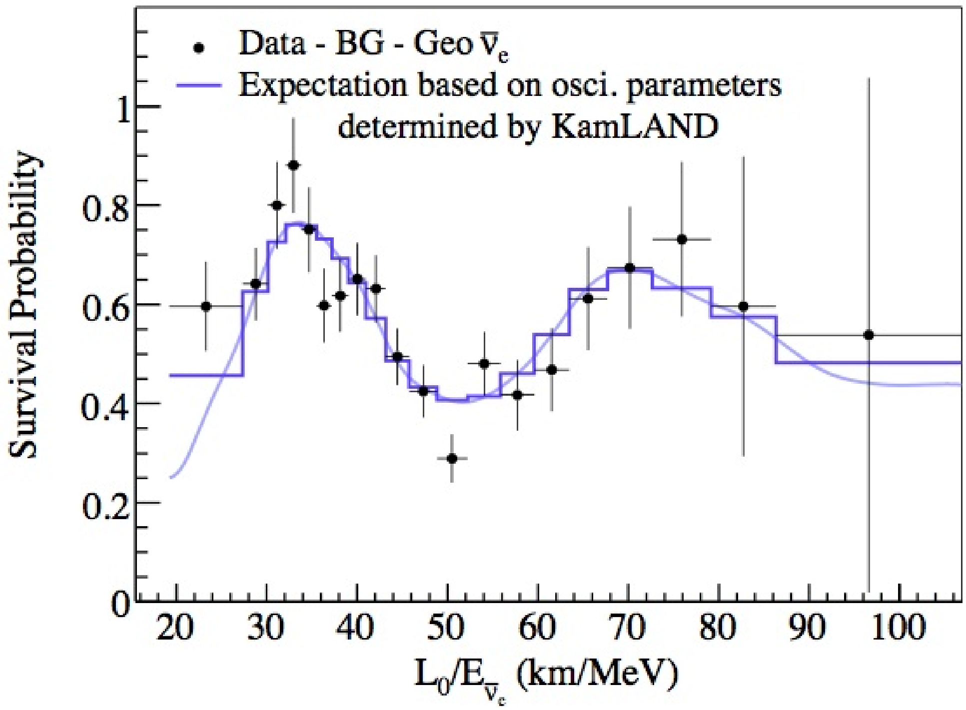


FIGURE 2. Neutrino Oscillation: Comparison of experimental data in Japan against that expected by theory. L is distance traveled in km. [4]

Aug 15, 2012: The ν_μ and $\bar{\nu_\mu}$ oscillation parameters do not indicate any new physics for anti-neutrinos. 2553 days of testing of atmospheric neutrino and anti-neutrino interactions at the MINOS Far Detector shows that the $|\Delta m^2| - |\Delta \bar{m}^2| \simeq 0.6 \times 10^{-3}\, eV^2$ difference is not yet statistically significant [25].

Since 1990, solar neutrino physics has evolved into a precision science. Data from the many neutrino experiments over the past decade is being assembled into a special 3x3 matrix checkerboard of key values. This is bit of a stretch to understand. Some of the neutrino oscillation data is in the form of mixing angles between the various types of neutrinos. We also need to know the tiny masses associated with the various neutrinos − but it is hard to get these directly. Neutrino conversion experiments are revealing Δm^2 differences not between e, mu, tau neutrinos themselves (called the three electroweak eigenstates) but rather their 'massive base states' labeled 1, 2, and 3 (called the mass-eigenstates). That is, neutrinos and antineutrinos are produced as ν_e, ν_μ, ν_τ together with the named charged leptons $\ell = e$, μ, τ. However, neutrinos of definite masses are something more primitive: ν_1, ν_2, ν_3. Similar to the 'CKM' matrix for quarks, these are connected by the 3x3 unitary transformation matrix and called the "PMNS" matrix (standing for Pontecorvo-Maki-Nakawaga-Sakata). One difference between this and the older CKM (Cabibbo, Kobayashi, Maskawa) quark matrix is that non-diagonal values are large (that is, wider mixings are more common). Like the neutrino case, the CKM quark matrix describes 'a Unitary rotation between the flavor eigenstates and the mass eigenstates.' PMNS shows the ability to mix generations or flavors. Again, its key working parameters are differences in mass-squared, 'mixing' angles, and also Dirac 'CP phase angle,' δ. The PMNS matrix form (there are others) just uses sines and cosines of mixing angles θ's and δ. For solar neutrinos involving loss of electron neutrinos, the 1 vs 2 difference is measured. For atmospheric muon neutrinos, the 2 vs 3 difference is measured. (For a little more detail, see Appendix at end).

Update 2014:
A 2014 PRL publication has the first demonstration at $> 5\sigma$ for the appearance of ν_e from ν_μ beam [39]. The values in the neutrino matrix (PMNS) are still not precise but are assisted by new measurements classified as "appearance measurements (e.g., muon neutrino beam results in the presence of electron neutrinos) and "disappearance (loss of flux of a particular neutrino state). Open questions still include the neutrino mass hierarchy and the value of the CP violating phase. Recent tests include a ν_μ beam from Tokai to Kamioka, Japan over 295 km distance (T2K). That is, from the 30 GeV protons at J-PARC to an off-axis pulsed ν_μ 0.6 GeV beam to the Super-Kamiokande (SK) 50 kT water Cherenkov detector. " A total of 28 electron neutrino events were detected with an energy distribution consistent with an appearance signal, corresponding to a significance of 7.3σ when compared to 4.92 ± 0.55 expected background events." However, the uncertainty still needs to be reduced further to pin down the CP violation parameter. SK discovered oscillation of atmospheric neutrinos in 1998, and T2K gave the first indication of flavor change in 2011.

Definitive results may have to await the completion of a new giant detector JUNO in China which should begin runs in 2019 [40] and data till 2025. The detector is a 38 meter diameter sphere of liquid scintillator below 700 meters of granite and will view reactor neutrinos from 53 km away.

3. Quark Gluon Plasma:

Lattice Quantum Chromodynamics calculations predict that matter should undergo a phase transition from the normal hadronic phase to a Quark-Gluon Plasma (GGC) phase at accumulated energy densities above 0.5 GeV/fm^3 (the volume of a proton is about two cubic fermi's). Present accelerators are now able to attain high energy densities near 15 GeV/fm^3 and hence should be seeing full QGC's. During the first few microseconds of the universe, the temperature was near 4 trillion degrees Celsius, and quarks and gluons existed as a deconfined plasma prior to forming protons and neutrons. The Brookhaven Relativistic Heavy Ion Collider, "RHIC," discovered in April 2005 that the experimental quark-gluon plasma (QGP) behaved as an unexpectedly perfect liquid without friction or viscosity [36]. High energy gold (Au) nuclei colliding on gold nuclei produce what is sometimes called the 'little big bang. Strange results continued to occur through 2010, and then the LHC did its own lead-on-lead Pb nuclei collisions in 2011 at twice the temperature (2.76 TeV per nucleon pair). This high temperature (trillions of degrees C) is still not yet hot enough to decompose the most tightly bound Upsilon particle (bottom, anti-bottom quarks) but does break apart their less tightly bound states. The J/ψ particle does break up (charm, anti-charm meson). Also the resulting back-to-back jet sprays are reduced on sides of the plasma with greater density − their energy is sapped by the plasma (this is called 'Jet Quenching). On the other hand, photons and Z bosons get through easily because they are not strongly interacting particles. Bottom quarks are heavier than charm quarks and tend to get through the plasma. With recent RHIC upgrades, it should be possible to distinguish between c and b quark particles.

On 3/10, the "STAR" detector at Brookhaven found the 'antihypertriton' (antiparticle nuclei of $p+n+\Lambda$ [quark structure 'sud']) with a lifetime of 2×10^{-10} seconds. They had previously seen anti-deuterium, anti-tritium, and anti-He-3. Strange quarks, s, are not rare in the quark-gluon plasma. Another observation is that the "fields created by gluons can twist, forming vortex-like structures in the all pervasive vacuum of space and when quarks loop through these vortices, they gain energy making them heavier." Off-center collision produce powerful magnetic fields causing charge separations with + charges moving in one direction and negative charges moving in another direction. The gluon created vortices are called "instantons." Recently, physicists in the RHIC/STAR collaboration observed that copper-copper collisions produce about 25% more strange quarks per nucleon than do gold-gold collisions [19]. Every new test is a learning experience. A future intention is to have uranium collisions, $U + U$.

A 'Glasma' (gluon plasma) is a hypothetical precoursor state of the QGP [26] [27]. When two relativistic nuclei collide, their Lorentz contraction makes them appear to be thin colliding disks (called color glass condensates). After these disks pass through each other, the space in-between becomes composed of highly coherent gluon coupled fields of high energy density called the glasma. It is this cylinder of energy that can then evolve into the QGP and eventually into a gas of ordinary hadrons (after 'cooling' down below about 150 MeV temperature.

4. Matter-AntiMatter Asymmetry (CP Violation):

A long-term interest for particle physicists has been why our universe is mainly made up of matter with very little antimatter. Three important operators in high energy physics are called 'C' for reversing the sign of charge, P for reversing parity, and T for time reversal. Parity refers to mirror image symmetry between left-ness and right-ness: should basic physics be the same in its mirror image? It turned out that the answer was, definitely "No!" Parity conservation was experimentally shown to be overthrown by weak interactions such as beta decay (Madame Wu, 1956) . In Feynman's view, an antiparticle is a particle moving backwards in time; and generally, basic physics looks the same under time reversal, T. It is still believed that the product of CPT operators is never violated (although tests are still ongoing to verify it). CP (charge parity) violation was first seen in 1964 in K-meson (kaon $K^o(d\bar{s})$) measurements (resulting in a Nobel prize for Cronin and Fitch in 1980). That in turn implies that T must also be violated meaning that the rate for a particle interaction is different for the time-reversed process (matter antimatter asymmetry).

After the discovery of the bottom quark, it was anticipated that CP violation would be much stronger in b-meson decays. Much data has now been gathered on B-factory tests with bottom-quark containing mesons like the B's. And indeed, measurement of large CP violation in the $B^o\,(d\bar{b})$ system was first observed in 2001 (at BABAR and Belle). Fermilab detectors in 2006 verified a mixing oscillation over time between B_s^o, $\bar{B}_s^o$ after long efforts (i.e., $s\bar{b}$, $b\bar{s}$ particle and antiparticle). CP violation in the decays of neutral 'charmed D-mesons' was seen by CERN in 2011 (the 'LHCb' experiment). LHCb made the first 5σ statistics observation of a CP asymmetry at the LHC in the mode $B_o \to K\pi$. The decays of bottom-mesons is a very lively arena awaiting a great many more publications.

The matter-antimatter oscillations in the charmed sector were the last to be observed. The $D^o(c\bar{u})$ can oscillate to the $\bar{D}^o(u\bar{c})$. Weak interactions create a slight mass difference between these two charmed mesons, and this affects their oscillation frequency and lifetimes. The physics revealed by the LHCb experiment involves determining the flavor of these mesons at production and then again at decay [37].

The 'Belle collaboration' of Japan ended in June 2010 after gathering short of a billion Upsilons ($b\bar{b}$, 4S) from electron-positron collisions. They studied bottom quark decays into charmonium (b $\to c\bar{c}s, : c\bar{c}d$, e.g., $B_o \to J$ + kaons or D+D-s [6]). Flavor changing neutral current radiative decays can also occur (b $\to s\gamma$). CP asymmetry is about 0.6%. The B $\to$ Charmonium K_o decays mediated by $b \to c\bar{c}s$ are experimentally clean and are called the 'golden modes' for seeing CP violation. The explanation for CP violation in the standard model is contained in a complex phase-angle in the CKM matrix describing quark mixing. But it probably comes from weak interactions rather than from QCD. The stronger cosmological matter-antimatter asymmetry probably depends mainly on some new physics beyond the Standard Model.

In late 2012, SLAC-BaBar finally and cleanly confirmed time asymmetry for the first time (and with 14 sigma certainty) [35]. T violation time asymmetry came from T-symmetry transformation and not CP in this case. The test used entangled neutral B^o-mesons from the decay of $\Upsilon(4S)$ [10.58 GeV from 9 GeV e^- and 3.1 GeV e^+] $b\bar{b}$ resonances formed from e^+e^- collisions and looked at the decay delay time for the CP odd and even output states. Ten years of data from 1999-2008 were processed with 5×10^8 $B\bar{B}$ mesons.

5. General:

The **strong force coupling constant** α_s is a major parameter of the standard model (SM). It has been generally claimed that its value is near unity for energies below a GeV. However, we know that α_s is actually not constant but rather decreases in value with energy (or momentum transfer, p_T). The decay in the curve is predicted by the "renormalization group equation" (and negative 'beta function'). The mass of the neutral Z-boson, $M_Z \simeq 91$ GeV, is a convenient reference energy to use for current particle physics, and at this energy $\alpha_s(M_Z) \simeq 0.116 < 1$ [13]. This reduced value of alpha can actually be calculated using both perturbative and latice QCD computations at short distances [29]. An up-to-date plot of the values of the strong coupling is shown in Figure 3. Why is this strength decay important? Recall that with high energies approaching the 'GUT' (Grand Unified Theory) scale, the strong, weak, and electromagnetic coupling constants are supposed to be comparable. We know that the electromagnetic coupling constant, α_{EM}, is weaker than the strong coupling but also increases in value when viewed at increasing energy (getting inside the electron-positron cloud surrounding an electron). Here the strong coupling gets weaker − making it easier to imagine that they might converge. The predicted convergence is supposed to be assisted by supersymmetry. However, there is some recent debate about being able to continue this sort of graph above the top quark mass ($M_t \simeq 173$ GeV). In addition, surprisingly, it might be the case that the curve declines again below about 200 MeV or so (low energy where peak $\alpha_s > 1.0$). This was recently modeled by Lattice-QCD [15]. Ongoing debate: − does computer modeling count as an experiment?
Some discussion of the Weak Force Coupling Constant is given in the Appendix at end.

The FermiLab Tevatron collider in Illinois began operation in 1985 but was permanently shut down on 9/30/2011. There will be no more TeV colliding proton beams in America. However, 'Fermilab' itself as an overall laboratory will continue to operate and will be doing important neutrino physics experiments (if adequate funding continues to be available).
Older high-energy collision data from Fermilab is still being analyzed. The CDF and D0 collaborations at the Tevatron experimentally discovered the top quark, t, in 1995. This is the most massive elementary particle known today (near 173 GeV) and couples very strongly with the Higgs boson, H. Because the Tevatron was a proton-antiproton collider, the $t\bar{t}$ resonances occur mainly by quark-antiquark annihiliation (center of mass energy near 2 TeV). The higher energy LHC is a pp collider so far near $\sqrt{s} = 7 - 8$ TeV center-of-mass energy giving it 22 times higher probability of $t\bar{t}$ production instead formed 85% by 'gluon fusion' $gg \to t\bar{t}$ [12]. In the standard mode, t decays nearly 100% of the time into a W and a b-quark. Gluon fusion is also believed to be a major way of producing the Higgs boson, $gg \to H$.

The LHC high energy 7 TeV inelastic pp 'cross section' implies a size of 0.86 fm (square, i.e., 73 mb [where 'b' = 'barn' = $10^{-24} cm^2$]) [4]. Total cross section rises with energy. Note

[4] "Big as a barn" for nuclear reactions. Integrated Luminosity is measured in inverse-femto-barns, fb^{-1} hinting at how many collisions occurred for a data base.

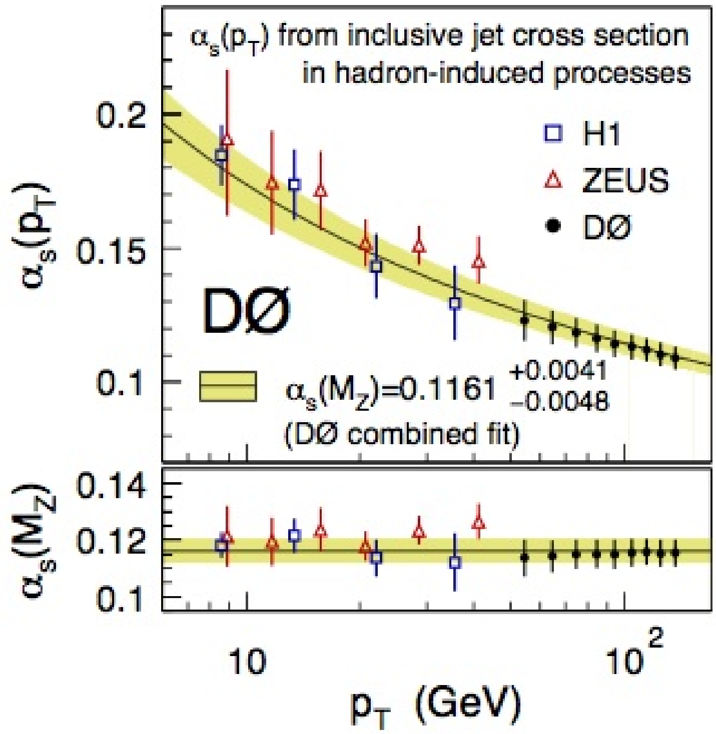


FIGURE 3. Experimental results for the declining coupling 'constant' α_s of the strong force versus energy. Earlier plots show $\alpha_s \simeq 0.4$ near 1 GeV to the far left.

that nuclear density is $n \sim 0.16\, fm^{-3}$ (or a square 1.84 fm − but proton diameter is 1.56 fm − intuitively not much wiggle room for motion of protons and neutrons − yet they do move fairly freely).

ATLAS (LHC, Oct-2011) has measured the probability of forming t-tbar (top +anti-top) but has not yet been able to distinguish single top quark production. Forward/Backwards motion asymmetry is not yet explained in CMS studies.

The status of "Weak Charge, Q_w:" Most fermions have a special non-electric charge that can contribute to the formation of weak bosons. This is measured by a parameter which is set approximately to minus one for the neutron, the source of natural beta decay, $Q_w(n) \simeq -1$. Although the B boson couples to fermions according to 'weak hypercharge' Y (where $Q = T_3 + Y/2$, and T_3 is 'weak-isospin'), the W and Z interact with anything that has non-zero Q_w. The weak vector bosons also carry weak charge and interact among themselves. Weak charge has chirality or 'handedness' that can lead to parity violation.

A useful formula for weak charge for nuclei is $Q_w(Z,N) \simeq Z(1-4\sin^2\theta_w) - N$, where theta is the Weinberg angle so that $x = \sin^2\theta_w \simeq 0.238$. [5] This is not actually a constant but varies somewhat with energy and with Z bosons versus W's. For Z's, it drops to near 0.23 near the mass of the bosons (the Z-pole). For the proton, $Q_w(p) = Q_w(1,0) \simeq 1(1-4(0.238)) = 0.048$. Since the square of the Weinberg angle is nearly one-fourth, the weak charge plays almost no role for protons. For the neutron however, $Q_w(n) = Q_w(0,1) \sim 0-1 = -1$, which has a large effect. These values have actually not yet been directly experimentally measured (the proton measurement is in progress). The 'Qweak' experiment at Jefferson Lab will measure the cross-sections for positive and negative helicity electrons in polarized elastic e-p scattering. There will be an asymmetry due to interference of photon and Z boson exchange. But large nuclei for cesium, thorium, bismuth, and lead have already been studied, and mesurements agree with SM and the formula and help pin down the Weinberg angle for those cases.

The most important measurement so far is the SLAC experiment 'E158' from 2003 in California. This is a measure of parity violation in Møller e-e scattering near 50 GeV using longitudinally polarized electrons scattering from unpolarized electrons in a 1.5 meter long liquid hydrogen target. The result is a left-right handed asymmetry near 0.14 ppm − small but definitely there and proportional to the weak charge. The weak charge for the electron now measured to be $Q_w(e) \simeq -0.04$. All of the quarks have weak charge with d being larger and positive while u is smaller and negative.

Tests of Left-Handedness of W bosons:

The weak W boson is a spin one particle so that its measured helicity can only take on values -1, 0, or +1. Unlike the spin-one photon, the finite mass of the W allows the existence of the zero polarization state and even encourages this so-called "longitudinal" polarization which is at right angles to the momentum vector. [6] The experimentally observed fractions of incidence of polarization are correspondingly called $f_-(\; or \; f_L)$, f_o, $and \; f_+$, and their sum must be 100%. The values depend on the test case and energy, but one would expect $f_L > f_R$ due to the left-handedness of the weak interactions. [7]

One of the latest tests for these incidences for the W boson come from the decay of the top quarks produced by the Fermilab high energy proton Tevatron. The top quark was actually discovered by CDF and D0 detectors at the Fermilab Tevatron in 1995, and the top quark chooses to decay to a W boson and a b quark nearly 100% of the time (tWb vertex of $t \to b + W^+$, and the b quark has to be left handed). [Note that charges go as $+2/3 \to -1/3 + 1$]. A "right handed polarization means that helicity is positive so that the spin vector of the W weak boson would point in the same direction as its momentum vector. This helicity choice is discouraged in the SM by $SU(2)_L$ weak forces. Although

[5] Here, 'Z' is total proton count, and N is neutron number in a given nucleus.

[6] For projections along the z axis, transverse polarizations would be given by $L_\pm = L_x \pm i L_y$ on L_z states.

[7] W interactions are purely LH "V-A", but Z interactions are not this simple and go like $(-ig_z/2)\gamma^\mu(c_v - c_A\gamma^5)$ where the vector and axial coefficients vary with each chosen fermion quark or lepton. $c_v = c_L + c_R = I^3{}_w - 2Q\sin^2\theta_w$.

near zero in this case, right polarization can still be as large as $f_+ \sim 3.6 \times 10^{-4}$ because of the relatively small mass of the b quark (on rare occasions, it can be found in a positive helicity state). The net average value of f_o turns out to be about 70% with experimental errors near $\pm 10\%$ [33]. The measured incidence f^+ is statistically near zero so that the remainder $f_- \sim 30\%$. [8] This left-handed helicity goes with the W^+ traveling away from the direction of top spin and away from the b motion, but the spin of the W is opposite to the direction of W momentum.

It is expected that with much higher energy, f_o will fall and f_L will increase (looking more like polarization of relatively massless particles). A recent example of this is 2011 results from the LHC- CMS Collaboration (thousands of authors[34]). In this case, the polarization of W bosons was studied for large transverse momenta in general pp collisions with the result that $f_o \sim 18\%$ and $f_L \sim 51\%$ (errors near $\pm 5\%$).

6. What Has Not Yet Happened (as of 2012):

Glueballs have not yet been found. The strong interaction mediated by gluons is 'non-abelian' meaning that gluons interact with other gluons. They should be able to get together to form a particle consisting of just glue, and there should be many mass states of these particles. "Nothing is more symbolic of the difficulty of solving QCD than the fact that, while glueballs are central to the understanding of non-perturbative QCD, there is currently no definite experimental evidence for their existence" [23]. This is largely a difficult signal-to-noise problem for experimenters.

A fourth neutrino has not yet been found [e.g., ν_s using ν_4 and m^2_{41}]. So, the mechanism of giving small masses to the neutrinos is still unknown (although there are a variety of possible 'seasaw' mechanisms − often mentioning right-handed neutrinos [all 'normal' neutrinos are left-handed]). There is also talk of a possible fourth generation of quarks, t' and b'. "The exploration of Terascale physics has only just started!"

QCD-Confinement: There is still no proof of confinement for quantum chromodynamics in the continuum limit (single quarks cannot escape from baryons) [15]. This problem is so difficult and so interesting that confinement is a Millennium Prize Problem from the Clay Mathematics Institute. How is it that massless Yang-Mills gluons enable ultimately massive bound states of gluons − the "mass gap. 'Establish rigorously the existence of the quantum Yang-Mills theory and a mass gap.' [Note that the short list of Millennium problems include the Poincaré Conjecture which was recently solved by Grigori Perelman].

SUSY: Repeated phrase from LHC publications: "No evidence is found for physics beyond the Standard Model." Supersymmetry is called SUSY, and its minimal supersymmetric extension onto the standard model is called 'MSSM.' Neither SUSY nor MSSM has yet

[8] In the relatively 'zero' b quark mass approximation, the value $f_o = {M_t}^2/({M_t}^2 + 2{M_w}^2) = 173^2/(173^2 + 2(80)^2) = 0.70-$ leaving $f_- \sim 0.30 = f_L = 2{M_w}^2/({M_t}^2 + 2{M_w}^2)$.

been found where it was supposed to be at the LHC − it was NOT just around the corner as some had previously claimed. One more recent example is an LHC CERN ATLAS summary of supersymmetry (SUSY) data which said, No excess above the Standard Model expectations is observed.. [8] [9] Exploration will have to continue now at higher energy. The Higgs mass near 126 GeV is difficult to achieve in the MSSM.

Nucleon Spin: There is still a "spin crisis" that only about 30% of nucleon spin is carried by quark spins. A lot of experimental and theoretical efforts for the last 20 years were devoted to search for the rest of the nucleon spin, without obvious success. One of the first observations that quark spin has low contribution came from the European Muon Collaboration of 1988 with subsequent verifications. "It is quite possible that much of the remaining nucleon spin will be found in the orbital motion of the valence quarks" [Jefferson Lab]. It is possible that "polarized glue" may contribute 50% of the proton's spin [31]. The pion cloud of a nucleon also contributes orbital angular momentum.

Cosmic Rays: "The mystery of the origin of cosmic rays is celebrating its 100th, anniversary in 2012"[14]. Charged cosmic rays should point toward their origin when their energy is $> 10^{20}\,eV-$ multiple EeV! (the highest energy so far detected). Energies beyond 10 PeV (peta $= P = 10^{15}$) are rare and are largely believed to originate within our galaxy from shock acceleration in supernova remnants. A variety of modern and special instruments is needed to cover over 8 orders of magnitude in energy and 24 in cosmic ray flux. One of the latest measurements (August, 2012) is proton-air-cross- section for particle interaction showers at incoming energy of 57 TeV (well above any current collider). The result is nearly 0.5 barns (25 times the geometric cross section of an incident proton) [24]. That also means that these ultra-high-energy cosmic rays won't make it to the surface of the earth. It is possible that IceCube has detected two neutrinos with energies above a PeV but it may take 7 years of total neutrino data to confirm this.

Dark Energy and Dark Matter have not yet been identified − and our WIMP experiments are not close to being able to pin down the nature of dark matter particles − if they indeed exist. And neither particle physics or astro-physics can succeed on its own − this is a joint venture.

References

[1] "Indication of Electron Neutrino Appearance from an Accelerator-produced Off-axis Muon Neutrino Beam," (Many Authors), http://arxiv.org/pdf/1106.2822v2
[2] Frank Close, Book Neutrino .
[3] http://en.wikipedia.org/wiki/Neutrino_oscillation

[9] The 'Natural SUSY' predictions were for super-partner 'higgsinos' near 200 GeV, 'stop/sbottom' near 600 GeV, and gluino's near 900 GeV [30].

[4] K. Nakamura and S.T. Petcov , Neutrino Mass, Mixing and Oscillations, May 2010, 53 pages review. http://pdg.lbl.gov/2011/reviews/rpp2011-rev-neutrino-mixing.pdf Picture is "kamland.jpg." Figure 13.4 from article.
[5] Brookhaven http://www.bnl.gov/rhic/news2/news.asp?a=1074& t=pr
[6] Himansu Bhusan Sahoo, Measurement of CP Violation in B Meson decays at Belle, hep-ex 1109.4780.
[7] arXiv:1108.3175 hep-ph 1 Oct 2011, from Beijing and Taiwan.
[8] arXiv:1110.0282, hep-ex 3 Oct 2011, Hideki Okawa. Matter e?ects and neutrino mass hierarchy.
[9] Carlo Giunti, ArXiv:1110.3914 18 Oct 2011, Phenomenology of Sterile Neutrinos.
[10] High-Energy Neutrino Astrophysics: Status and Perspectives U.F. Katza , Ch. Spieringb, arXiv: 1111.0507 long.
[11] Bhattacharya, et. al., "Implications of sterile neutrinos for medium/long-baseline neutrino experiments and the determination of Theta 13." , ArXiv: 1111.4225 hep-ph 17 Nov 2011.
[12] Frederic Deliot, "Top Quark Physics at Hadron Colliders," arXiv:1111.6274 hep-ex 27 Nov 2011.
[13] Markus Wobisch, "Recent QCD results from the Tevatron, Nuclear Physics B Proceedings Supplement 00 (2012) 112, http://arxiv.org/pdf/1202.0205.pdf hep-ex 1 Feb 2012.
[14] Angela V. Olinto, "AstroParticle Physics at the Highest Energies, 32-nd International Cosmic Ray Converence, Beijing 2011, http://arxiv.org/pdf/1202.0355.pd
[15] Ph. Boucaud, et. al., "The Infrared Behaviour of the Pure Yang-Mills Green Functions," hep-ph 9 Sep 2011 (50 pg. review, see important plot p. 21). http://arxiv.org/pdf/1109.1936v1.pdf.
[16] NIST 2010 "CODATA International recommended values of the Fundamental Physical Constants," http://physics.nist.gov/cuu/Constants/index.html
[17] The Weak Interaction, 4/14/2011, http://www2.warwick.ac.uk/fac/sci/physics/teach/module_home/px435/weak.pdf
[18] D.M. Webber, "Toward a sub-ppm measurement of the Fermi constant,"arXiv hep-ex 1006.3982 21 Jun 2010.
[19] "Copper collisions create much strangeness," http://physicsworld.com/cws/article/news/48295, 1/13/2012.
[20] "Observation of electron-antineutrino disappearance at Daya Bay, ArXiv hep-ex 1203.1669 8 Mar 2012, also see: Zhi-zhong Xing (Beijing) "Implications of the Daya Bay observation of θ_{13} on the leptonic flavor mixing structure and CP violation," ArXiv 1203.1672 12 Mar 2012.
[21] Stefan Antusch, et. al., "$\theta_{13}^{PMNS} = \theta_c/\sqrt{2}$ from GUTs," arXiv:1205.1051 [hep-ph] 9 Aug 2012.
[22] Matt Strassler's web site for: Of Particular Significance, Conversations about science."If the Higgs Field were Zero" http://profmattstrassler.com/articles-and-posts/particle-physics-basics/the-known-apparently-elementary-particles/the-known-particles-if-the-higgs-field-were-zero/
[23] E. Gregory, et. al., "Toward the glueball spectrum from unquenched lattice QCD," for JHEP, arXiv:1208.1858 [hep-lat] 9 Aug 2012.
[24] Pierre Auger Collaboration (1.5 pages of authors), "Measurement of the proton-air-cross-section at 57 TeV with the Pierre Auger Observatory," ArXiv: [hep-ex] 1208.1520 9 Aug 2012 (results agree with extrapolation of LHC energy).
[25] Fermilab, "Measurements of atmospheric neutrinos and antineutrinos in the MINOS Far Detector," arXiv:1208.2915v2 [hep-ex] 15 Aug 2012.
[26] Larry McLerran, "The CGC and the Glasma: Two Lectures at the Yukawa Institute," arXiv:1011.3204v1 [hep-ph] 14 Nov 2010.
[27] B.V. Jacak, B.Müller, "The Exploration of Hot Nuclear Matter," Science Vol 337, 20 July 2012, p 310.
[28] Lisa Grossman, "New particle, new questions," New Scientist 14 July 2012 p6.
[29] Xavier Garcia i Tormo, "α_s from the static energy in QCD," arXiv:1208.4850 [hep-ph] 23 Aug 2012.
[30] G. Redlinger, "Searches for Supersymmetry with the ATLAS Detector," arXiv:1209.1742 [hep-ex] 8 Sep 2012.
[31] C. Aidala, et al., "The Spin Structure of the Nucleon," arXiv:1209.2803 [hep-ph] 13 Sep 2012. Also see a previous article on missing spin: ArXiv 1111.2562.

[32] Johanna Miller, "The Higgs particle, or something much like it, has been spotted," Physics Today, September 2012, pp. 12-15.
[33] B. D. Gmyrek, Dissertation, 2007, "Measurement of the W Boson Helicity in Top Quark Decays," http://www-d0.fnal.gov/results/publications_talksthesis/gmyrek/thesis.pdf
[34] CMS Collaboration (thousands of authors), "Measurement of the Polarization of W bosons with Large Transverse Momenta in W^+ Jets Events at the LHC," arXiv:1104.3829v1 [hep-ex] 19 Apr 2011.
[35] The BaBar Collaboration, "Observation of Time-Reversal Violation in the B^o Meson System," Phys. Rev. Lett. 109, 211801 (2012)
[36] M. Tannenbaum, "Highlights from RHIC plus some earlier BNL highlights in subnuclear physics," ArXiv 1211.4527
[37] R. Aaij et al. (LHCb Collaboration), "Observation of $D^o - \bar{D}^o$ Oscillations," Phys. Rev. Lett. 110, 101802 (2013) March 5, 2013 PDF free, see http://physics.aps.org/articles/v6/26 .
[38] L.J. Hall and G.G. Ross, "Stitching the Yukawa Quilt in light of Θ_{13}, arXiv:1303.6962, CERN.
[39] Joseph A. Formaggio, "Viewpoint: Neutrino Experiments Come Closer to Seeing CP Violation," http://physics.aps.org/articles/print/v7/15, and see ref: (many authors) "Observation of Electron Neutrino Appearance in a Muon Neutrino Beam," PRL 112, 061802 (2014).
[40] Jane Qiu "China Builds Mammoth Detector to probe Mysteries of Neutrino Mass," Science, 343, 7 Feb 2014 p 590-591.

7. Appendix:

More on the Neutrino Matrix:

Some neutrino oscillation parameters are beginning to be pinned down [4] [21]. A current list of known values is:

$$\Delta m^2_{12} \simeq 7.6 \times 10^{-5}\, eV^2, \quad \theta_{12} = \theta_{Solar} \simeq 34.0^o \pm 1.1^o(1\sigma)\ .$$
$$|\Delta m^2_{23}| \simeq 2.4 \times 10^{-3}\, eV^2 \quad \theta_{23} = \theta_{Atmospheric} \simeq 46.1 \pm 3.4^o(1\sigma).$$
$$\Delta m^2_{13} \simeq 2 \times 10^{-3} eV^2 \qquad \theta_{13} = 8.8 \pm 1.0^o(1\sigma) \quad > 0!\ ^{10}\ .$$

$\Delta m^2_{32} \simeq \Delta m^2_{31}$, and $|\Delta m^2_{31} - \Delta m^2_{32}| = |\Delta m^2_{21}|$. Presently, the sign of Δm^2_{atm} is unknown. The (3,2) neutrino mixing values are doubly determined independently by atmospheric and by accelerator experiments. The (2,3) values are also doubly determined by solar and KamLAND experiments. Now (1,3) is also double covered by reactor and by accelerator experiments. A current puzzle is that we do not yet know the masses of the base states (eigenstates): ν_1, ν_2, ν_3 [8]. We know that $m_2 > m_1$ but don't know if m_3 is larger than these or smaller (heirarchy problem). It is expected that eventually double-beta decay experiments may provide the answer − in case neutrinos are 'Majorana' particles (their own antiparticle). Equally promising are long-baseline neutrino accelerator experiments, provided $\sin 2\theta_{13} \sim 0.001$. Also a 100 Megaton detector for neutrinos may give the answer if $\sin^2 2\theta_{13} > 0$. The optimal test length for θ_{13} is L = 0.5 km E/MeV; so do 1-2 km short range testing. The 'Chooz' reactor in France used 1 km, and 'Double Chooz' is next.

[10] A March, 2013 summary over five reactor experiments gives $\Theta_{13} \sim 13^o-$ a higher value [38]. This very interesting lack of suppression is much stronger than the CKM case ($\theta_{13} \simeq 0.2^o$).

As a simple example of a 2x2 subset of 3x3 matrix, consider the case of two neutrino avours ν_μ, ν_τ and two mass eigenstates ν_2, ν_3. One has a superposition of states:
$\nu_\mu = \nu_2 \cos\theta_{23} + \nu_3 \sin\theta_{23}; \quad \nu_\tau = -\nu_2 \sin\theta_{23} + \nu_3 \cos\theta_{23}$.

If the masses m_2 and m_3 are dfferent, quantum mechanical time evolution of an initial ν_μ state induces a non-zero transition probability to ν_τ. The survival probability for the muon neutrino is:

$$P(\nu_\mu \to \nu_\mu) = 1 - \sin^2 2\theta_{23} \sin^2[1.27\Delta m^2_{23}\, L[GeV]/E_\nu[eV^2\, km]]. \tag{1}$$

where L (in km) is the distance travelled by the neutrino, E_ν (in GeV) its energy, and $\Delta m^2_{32} = m^2_3 - m^2_2$ (in eV^2). Notice that the division by energy means that the oscillation is fast and wild at low energy, but most testing is done in the GeV's range.

The present estimation of the PMNS Unitary matrix is: [11]

$$|U| = \begin{vmatrix} 0.8-0.9 & 0.5-0.6 & 0.0-0.2 \\ 0.3-0.6 & 0.3-0.7 & 0.6-0.8 \\ 0.1-0.5 & 0.5-0.8 & 0.6-0.8 \end{vmatrix} \tag{2}$$

If there are neutrinos lying beyond the basic three, then their masses are constrained by cosmological requirements to sum to less than a total of 0.6 eV.

<u>The Weak Coupling Constant:</u>

The coupling constant for the weak force is often presented as the historically old 1932 Fermi Coupling, G_F which has the tiny value $G_F/(\hbar c)^3 = 1.166 \times 10^{-5} GeV^{-2}$ [16]. In 2010, this value was measured to better than a part-per-million accuracy based on the mean life of positive muons $\tau \simeq 2.197 \mu s$ [18]. When people say that the weak force is about a million times weaker than the strong force, they are referring to this Fermi constant. But another more currently relevant form is $\alpha_w = {g_w}^2/4\pi \simeq 1/30$. The g-coupling is attached to each vertex of the Weak exchange Feynman diagram; and $g_w \simeq 0.65$ is related to the mass of the charged W vector boson, $m_w \simeq 80.4$ GeV (the mass itself is contained in what is called the propagator).

$$\frac{G_F}{(\hbar c)^3} = \frac{\sqrt{2}\, g^2}{8{m_w}^2} \tag{3}$$

Since the electromagnetic coupling $\alpha_{EM} \simeq 1/137$, we note that α_w is in fact nearly four times stronger than EM! The weakness of the weak interaction is due to its having a low probability of occurrence which in turn is due to the large mass of the relevant W boson. And, at high energies where momentum transfer is near the W mass, then the weak interaction is comparable in strength to the electromagnetic interaction [17].

How do the coupling constants for the basic forces change and begin to converge in value with increased energy? There is a basic concept for Renormalization Group Equations (RGE) called the important 'beta function (or Yang Mills single loop beta function) which indicates the slope of coupling constant curves with energy. [11] This leads to expressions like:

$$1/\alpha(q^2) = 1/\alpha(\mu^2) - \beta \ln(q^2/\mu^2). \tag{4}$$

μ *or* m is a reference value (e.g., 1 GeV energy scale for QCD) and q is the momentum transfer in an experiment. If beta is a constant, then 'inverse coupling' constants will scale with momentum transfer as Straight Line Plots. 't Hooft surprisingly found in 1973 that beta for QCD was negative! so that a graph of 'inverse coupling' versus log(energy) will rise strongly towards higher energies (and give asymptotic freedom!). A frequent choice is $\beta = [2n_f - 11n_b]/12\pi = SF/12\pi$ for ['slope factor'].
Case 1 EM: Photons do not generate other photons, so useful $n_b = 0$ and $\beta = +1/\pi > 0$ (no slope, just log energy dependence, and $n_f = 6$) .
2) For Weak $SU(2):$ $n_b = 2$, so $\beta = [12 - 22]/12\pi = -5/6\pi < 0$. $n_b = N$ of SU(N).
3) For Strong Color $SU(3)$, $n_b = 3$, $\beta = [12 - 33]/12\pi = -7/4\pi < 0$. [12]
The coupling constants still will not converge to each other at ultra high energies unless the basic theory is beyond the SM (e.g., MSSM super-symmetry).

Note that although Lattice computations are fairly successful for QCD and the electromagnetic force, "the difficulties in simulating the weak nuclear force and gravity on a lattice have so far proved insurmountable."

---

[11] the beta slope is defined by: $\beta = \mu\partial g/\partial\mu = \partial g/\partial \ln \mu$.

[12] Plugging these values into the above equation will only crudely duplicate the strong coupling plot and the changes in the electromagnetic α_{EM} coupling. Reality may be slightly more involved. The beta function is really a series with $\beta \simeq \beta_0$ only being an approximation for one-loop and higher powers of α for other terms (β_1 for 2-loops, β_2 for 3-loops). It is also possible to consider the number of active quarks as incrementing by 1 when μ crosses a quark mass threshold m_f. So, projecting α_s back from m_Z would use 5 quarks until m_b is attained at 4.2 GeV and another integer below m_c at 1.27 GeV.

ROTATIONS OF BASE STATES

DP

ABSTRACT. Some of the elementary particle states which we might call 'basic' fermion flavors have to be rotated into mixed states to function in the world of weak interactions. For example, the weak interaction for quarks can be said to depend on a rotated d-prime: $d' = d \cos\theta_c + s \sin\theta_c$ where θ_c is the Cabibbo angle [1]. The familiar intermediate vector bosons are also combinations and rotations of basic gauge boson states, and now the neutrinos seem to be combinations of basic states with tiny masses causing them to interconvert into each other's flavor states.

1. INTRODUCTION:

Gluons interact with quarks, and weak bosons interact with leptons according to familiar flavor names for elementary fermion particles. But, physical weak bosons see quarks in a strange or "twisted" way using base states rotated into mixtures. The reason for the particular choice selected by Nature is not known, but it is phenomenologically treated by stating new sets of parameters within the Standard Model. In 1963, Nicola Cabibbo proposed a quark doublet of u, d' to account for observed weak decays of strange particles and allow for the mixing of the d and s quarks with a mixing angle near $\theta_c \simeq 13^o$. It is useful to treat the u, d' as a 'doublet' and the c, s' also form a doublet [1].

Forming doublets within flavor generations is basic for weak interactions, and flavor changing transitions within doublets are preferred [2]. The charm quark, c, was introduced theoretically in 1970 by Glashow, Iliopoulos and Maiani (the "GIM mechanism") to cancel out an unwanted $\Delta S = 1$ strangeness changing neutral current [3]; that is, charm was suggested to keep the strange quark from doing something. The term "charged currents" means that charged leptons are produced, while "uncharged" means the absence of charged leptons (e, μ, τ). In the Cabibbo model with only three quarks (u, d, s), it was very hard to explain the absence of $\Delta s = 1$ neutral currents (such as K^+ decays without μ's). Verification of the strange GIM idea came in 1974 with the discovery of the $J/\psi = c\bar{c}$ meson. With the later discoveries of bottom and top quarks (b and t), the overall mixing superpositions required a new unitary 3×3 matrix $\mathbf{V}$ called the CKM matrix (Cabibbo, combined with

Date: November 20, 2011.

email: davepeterson137@gmail.com. Paper updated to May 12, 2014.

[1] In the Weinberg-Salam model, we have left-handed doublets such as $\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$. And then for quarks, suggest $\begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L$, with $s' = -d \sin\theta_c + s \cos\theta_c$.

Kobayashi and Maskawa, 1972) [2].

In the realm of leptons, neutrinos are only able to mix together if they possess mass − even a tiny mass. We now know that they do, and we are able to detect some of the mixing between families of leptons. Like CKM, this set of mixing values is an undetermined parameter set added to the standard model of particle physics.

2. Quark Flavor Changes and the CKM Matrix:

In the mid-20th century, it was noted that when particle decays include leptons (e, μ, ν's), the decay rates with strangeness changing $\Delta S = 1$ are suppressed below those with $\Delta S = 0$ by a factor of about twenty, e.g,

$$\frac{P(s \to u)}{P(d \to u)} \simeq \frac{(G \sin\theta_c)^2}{(G \cos\theta_c)^2} = (\tan\theta_c)^2 \simeq (\tan 13^o)^2 \simeq 1/20. \tag{1}$$

where G is the weak Fermi coupling constant representing pure lepton decays like $\mu \to e + \nu$'s. That is, for flavor changes, a d quark prefers to decay to a u quark, but an s quark decaying to a u quark is more suppressed. An example of this is pion versus kaon decay: e.g.,

$$\pi^- = d\bar{u} \to u\bar{u} + (W^- \to \mu^- + \bar{\nu}_\mu), \quad K^- = s\bar{u} \to u\bar{u} + (W^- \to \mu^- + \bar{\nu}_\mu). \tag{2}$$

In the pre-charm world of only three quarks (u,d,s), Nicola Cabibbo declared that the weak interaction charged current should look like:

$$J_\mu(x) = \bar{u}(x)\gamma_\mu(1+\gamma_5)[\cos\theta_c\, d(x) + \sin\theta_c\, s(x)] \tag{3}$$

That is, the u-quark couples with a combination of d and s quarks now called d'. This means, for example, that the neutron (udd) decay coupling constant is slightly less than the muon decay constant, $G\cos\theta_c < G$. The Cabibbo angle represents a strong-versus-weak flavor symmetry breaking direction using a convention to place the quark mixings into the down type quarks (d',s',b'). It is a general convention to leave the "u" type quarks in the upper portions of the doublets intact (u,c,t). The identity of the base states in the strong force Lagrangian is different from those of the electroweak Lagrangian using d', s', b'– superpositions of mass eigenstates. This concept is not explained by the Standard Model (SM) and is an added undetermined parameter. In the Standard Model, the 'physical' quark states or mass eigenstates do not act as pure states in weak interactions. The first unusual early requirement of weak interactions was that they have '$V - A' \simeq$ a $\gamma_\mu(1-\gamma_5)$ term [3] in Fermi type interactions (standing for vector minus axial-vector contributions). The addition of a symmetry breaking direction is beyond just $V - A$ and arises somehow from Yukawa interactions with the background Higgs condensate [5]. Finally , CKM (1973)

---

[2]And Nobel Prize in Physics in 2008 "For the discovery of the origin of the broken symmetry which predicts the existence of at least three families of quarks in nature" − shared in addition with Yoichiro Nambu for his ideas on broken symmetry

[3]The gamma's refer to Dirac matrices in which γ_μ's behave like vectors and $\gamma_\mu\gamma_5$ behaves like an axial vector flipped under mirror symmetry.

improved this equation to include $(\bar{u}, \bar{c}, \bar{t})$ coupled with the rotated (d', s', b') times the weak W_μ's.

So, the CKM V matrix produces rotated superposition states like this:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{4}$$

The present best values for the CKM Unitary matrix as (u,c,t) by (d,s,b) quarks are:

$$V_{CKM} = \begin{vmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{vmatrix} = \begin{vmatrix} \underline{0.974} & 0.225 & 0.0035 \\ 0.225 & \underline{0.9734} & 0.041 \\ 0.0086 & 0.040 & \underline{0.99915} \end{vmatrix} \tag{5}$$

Unlike the neutrino mixing matrix, notice that the the diagonal elements of this CKM matrix are close to one in value. There are many different parameterizations of the CKM matrix. The original KM version used three angles and a CP violating phase angle $(\theta_1, \theta_2, \theta_3, \delta)$. The matrix element $V_{11} = V_{ud} = c_1 = \cos\theta_1$ for example where $\theta_1 = \theta_c$ is the Cabibbo angle. Now a more 'standard' choice parameterization uses three Euler angles $(\theta_{12}, \theta_{23}, \theta_{13}, \delta_{13}$ where $\theta_{12} = \theta_c = 13.04^o$ is still the Cabibbo angle, and $V_{12} = V_{us} = c_{12}c_{13} = \cos\theta_{12}\cos\theta_{13} \sim \cos\theta_c$. In this case, $\theta_{13} = 0.20^o, \theta_{23} = 2.4^o, \delta_{13} \simeq 1.2^o$ [5]. A picture of this rotation of base states using just the Cabibbo angle is shown in [7]. These parameter sets do not strictly derive from SM and are added undetermined parameters of presently unknown origin. An example of the use of the CKM matrix is this equation:

$$i\,(= u, c, t) \to Vertex\, V_{ij}\sqrt{G_F} \to W^* + j\,(= d, s, b) \tag{6}$$

Since diagonal elements of CKM are near unity, this again says that u has a preference to interact with d, c with s, and t with b − staying within generations.

Quarks are confined within hadrons. The masses of the lightest quarks are hidden by the gluon fields of hadrons and can only be found by deductions consistent with lattice quantum-chromodynamics (L-QCD) numerical modeling calculations to be near: $m_u = 2.01 \pm 0.14$ MeV, $m_d = 4.79 \pm 0.16$ MeV, and $m_s = 92.4 \pm 2.5$ MeV [4].

3. Weinberg Bosons:

The ElectroWeak theory is often labeled by its 'unified' Lie group symmetry: $SU(2) \times U(1)$ [actually $SU(2)_L \times U(1)_Y$ for left-handed leptons and Y-hypercharge − a combination

[4] Wikipedia says 95 ± 5 MeV. Masses for quarks are only deduced by total self-consistencies of lattice QCD calculations which in turn have to agree with the constraints of phenomena. So the term mass-basis for quarks doesn't make much sense except as physical flavors. Weak interactions just see them through twisted (or Picasso broken) glasses. Quarks are bound and never have long Feynman diagram legs.

of electric charge and isotopic spin 'charge']. In electromagnetism, the interaction of electric current with an EM field vector potential is given by $eJ_\mu A_\mu$ where $A_\mu = (\phi, \vec{A})$ = 4-vector potential from which EM fields E and B can be calculated. Somewhat analogously, the Weinberg-Salam model uses an electroweak interaction energy density with Lagrangian terms $L = gJ_\mu \cdot W_\mu + g'(J^{em}_\mu - J^{(3)}_\mu)B_\mu$ [3].
Initially,there are four massless mediating gauge boson fields − three from SU(2) and one isoscalar from U(1): $W_\mu = (W^{(1)}_\mu, W^{(2)}_\mu, W^{(3)}_\mu)$ and, B_μ. After spontaneous symmetry breaking using the Higgs field, massive charged vector bosons are formed as W^+, W^-:

$$W^{\pm}_\mu = \frac{W^{(1)}_\mu \pm iW^{(2)}_\mu}{\sqrt{2}} \tag{7}$$

And there is a neutral massive Z boson and a massless photon A_μ formed by mixing of neutral fields:

$$A_\mu = B_\mu \cos\theta_W + W^{(3)}_\mu \sin\theta_W, \quad Z_\mu = -B_\mu \sin\theta_W + W^{(3)}_\mu \cos\theta_W \tag{8}$$

or in terms of an explicit matrix rotation of the neutral fields:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^{(3)}_\mu \end{pmatrix} \tag{9}$$

where θ_W is the Weinberg or Glashow or weak mixing angle $\simeq 28.9^o$. This means that the electromagnetic times weak symmetry group combination $U(1) \times SU(2)$ isn't pure but is somewhat kludged together using the Weinberg angle. The coupling constants are related by $g'/g = \tan\theta_W$ and $e = g\sin\theta_W = g'\cos\theta_W$. The masses of the intermediate vector bosons depend on the vacuum expectation value of the Higgs field. But $M_W = M_Z \cos\theta_W$.

4. Neutrino Mixing:

Neutrino conversion experiments are revealing Δm^2 differences not between e, mu, tau neutrinos themselves (called the three electroweak flavor eigenstates) but rather their "massive" base states 1,2,3 (called the mass-eigenstates). That is, neutrinos and antineutrinos are produced as ν_e, ν_μ, ν_τ together with the named charged leptons $\ell = e$, μ, τ. However, neutrinos of definite masses are something more primitive: ν_1, ν_2, ν_3. These are connected by the 3x3 unitary transformation similar to the CKM matrix for quarks and now called the "PMNS" matrix (standing for Pontecorvo-Maki-Nakawaga-Sakata). The origin of non-zero neutrino masses comes from something beyond the current Standard Model (as also does Dark Matter).

Key working parameters are differences in mass-squared, 'mixing' angles, and Dirac CP phase angle, δ. The PMNS matrix just uses the θ's and δ. For solar neutrinos involving loss of electron neutrinos, the 1 vs 2 difference is measured. For atmospheric muon neutrinos, the 2 vs 3 difference is measured. Some neutrino oscillation parameters are beginning to be pinned down [6]. A current list of known values is:

$$\begin{array}{ll} \Delta m_{12}^2 \simeq 7.6 \times 10^{-5}\, eV^2, & \theta_{12} = \theta_{Solar} \simeq 34^o, \text{ range: } (30^o - 38^o). \\ |\Delta m_{23}^2| \simeq 2.4 \times 10^{-3}\, eV^2, & \theta_{23} = \theta_{Atmospheric} \simeq 45^o. \text{ range: } (37^o - 56^o). \\ \Delta m_{13}^2 \simeq 2 \times 10^{-3} eV^2, & \theta_{13} \text{ range: } 0^o - 13^o. \text{ (now 9 degrees, 3/12).} \end{array}$$

The mixing of neutrino states is shown by the U_{PMNS} Mixing Matrix:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{vmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{vmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \tag{10}$$

The present estimation of the PMNS Unitary matrix is not precisely known but is limited in ranges to:

$$|U| = \begin{vmatrix} 0.8 - 0.9 & 0.5 - 0.6 & 0.0 - 0.2 \\ 0.3 - 0.6 & 0.3 - 0.7 & 0.6 - 0.8 \\ 0.1 - 0.5 & 0.5 - 0.8 & 0.6 - 0.8 \end{vmatrix} \tag{11}$$

There are now many beginning and ongoing neutrino experiments across the globe. They should contribute to the 'neutrino revolution' at a rapid pace. Revealing the next model beyond the current Standard Model (e.g., SO(10)) should clarify the origin of neutrino mass.

March 2012: "The Daya Bay Reactor Neutrino Experiment has measured a non-zero value for the neutrino mixing angle θ_{13} with a significance of 5.2 standard deviations [8] using antineutrinos from six reactors with 55 days of data. The result is $\theta_{13} \simeq 8.8 \pm 0.8^o (\pm 1\sigma$ range)[5]. This very interesting lack of suppression is much stronger than the CKM case ($\theta_{13} \simeq 0.2^o$). Previous experiments from T2K, MINOS, and Double Chooz had hinted at a non-zero PMNS-matrix angle. This matrix 'describes a fundamental mismatch between the weak-interaction (flavor) and mass eigenstates of six leptons.' In the CKM case, there was a relationship between angles and the strong hierarchies of quark masses (e.g., $\sin\theta_c \simeq \sqrt{m_d/m_s} = \sqrt{4.79MeV/92.4MeV}$) $[\theta_c \sim 13.1^o]$. A similar analogy for leptons might say $\sin\theta_{23} \simeq \sqrt{m_\mu/m_\tau} + \sqrt{m_2/m_3} \simeq 0.65-$ actual result $\theta_{23} \simeq 45^o$.

References

[1] Francis Halzen, Alan D. Martin, Quarks and Leptons, An introductory course in modern particle physics, 1984, John Wiley.

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

[3] Donald H. Perkins, Introduction to High Energy Physics, 4th Ed., 2000, Cambridge.

[4] Michio Kaku, Quantum Field Theory, A modern introduction, 1993, Oxford.

[5] A. Ceccucci, et. al., The CKM Quark-Mixing Matrix, Wikipedia.org, Feb 2010,

[5] A March, 2013 summary over five reactor experiments gives $\Theta_{13} \sim 13^o-$ a higher value [9].

[6] K. Nakamura and S.T. Petcov , Neutrino Mass, Mixing and Oscillations, May 2010, 53 pages review. http://pdg.lbl.gov/2011/reviews/rpp2011-rev-neutrino-mixing.pdf Picture is "kamland.jpg." Figure 13.4 from article.

[7] Martinus Veltman, Facts and Mysteries in Elementary particle physics,World Scientific, 2003 (see pg. 104, 106).

[8] "Observation of electron-antineutrino disappearance at Daya Bay, ArXiv hep-ex 1203.1669 8 Mar 2012, also see: Zhi-zhong Xing (Beijing) "Implications of the Daya Bay observation of θ_{13} on the leptonic flavor mixing structure and CP violation, ArXiv 1203.1672 12 Mar 2012.

[9] L.J. Hall and G.G. Ross, "Stitching the Yukawa Quilt in light of Θ_{13}, arXiv:1303.6962, CERN.

THE FINE STRUCTURE CONSTANT

DP

1. Introduction:

Feynman presented his diagrams in 1948 as a mneomonic aid for writing down appropriate integrals for calculations. After clarification by Freeman Dyson, they finally caught on in a big way and now appear in many thousands of papers and on black(white) boards in universities helping to to simplify fairly complex ideas. Gell-Mann was a little irritated by Feynman and always called his diagrams 'Stueckelberg diagrams' after a Swiss physicist who had used them earlier than Feynman. A Feynman diagram represents a quantum field theory (QFT) process in terms of visualizable particle paths. Feynman went further than his predecessors and had more useful ideas. His diagrams are are a composed of a collection of vertices connected by external and internal lines (e.g., see Figure 1 below). Each vertex is an intersection of two fermions and an interacting boson. The example shown is an external electron in, and electron out, and a photon as the interaction boson (an integral spin 'particle').

This note is concerned mainly with just the coupling constant attached to each vertex of a Feynman diagram. For electromagnetism, the strength of a vertex depends on the "fine structure constant," $\alpha \simeq 1/137$ being proportional to the square of the charge of the electron e^2. There is a similar but much larger constant for the strong interactions, $\alpha_s \sim 0.12$. A popular convention is to attach a value of $g_e = \sqrt{4\pi\alpha}$ to each interaction point of these diagrams. Is there a simple way to explain this choice without refering to the difficult mathematics of quantum field theory? Yes. Like first order QFT results, sample classical scattering calculations also result strengths proportional to e^4. In quantum mechanics (QM), these results come from the 'Born rule' that classical probabilities are QM 'amplitudes' squared; and each amplitude is the result of the emission and the absorption of a photon– two steps. So each vertex step must have a strength $\sqrt{\ } of \sqrt{\ } of\ e^4$ or just $e \propto \sqrt{\alpha}$. Where possible, quantum mechanics has to agree with classical results $-$ the 'Bohr correspondence principle.'

Note that these constants α_e, α_s are not quite really constant, they change slightly with the high energy that tries to see them. At high est energies now available, for example, the value of α may change from 1/137 to 1/128. This is due to the electron being surrounded

Date: November 18, 2011.

email: davepeterson137@gmail.com. Paper updated to November 20, 2011.

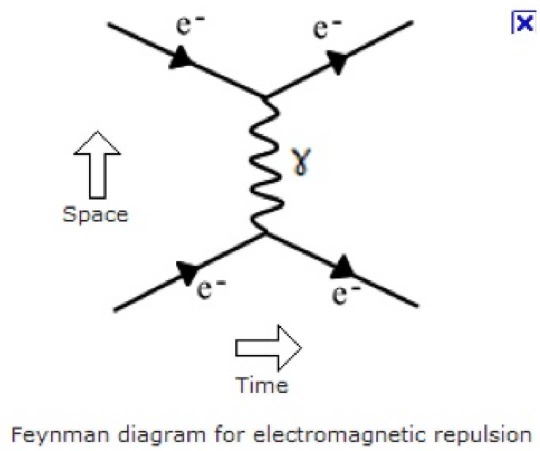


FIGURE 1. The simplest Feynman Diagram picture. Label the vertices at the ends of the vertical virtual photon wiggly line with $g_e = \sqrt{4\pi\alpha}$. People often invert the directions of space and time so that time runs vertically. Here, two e's approach each other from left to right, feel a repulsive force, and then diverge.

by a cloud of virtual electron -positron pairs which shields its true strength. A high energy probe can see part-way inside that cloud to observe a different effective charge.

2. Value of alpha:

The fine structure constant of the electron can be expressed in several different ways

$$\alpha = \frac{e^2}{4\pi\epsilon_o\hbar c} \quad or\ just \quad = \frac{e^2}{\hbar c} \quad or \quad = \frac{e^2}{4\pi} \tag{1}$$

depending on which system of units is being used ['SI,' 'ESU,' 'Natural,' ...]; there are many systems of units in use by physicists. These forms all have in common that the constant, α is proportional to the square of the charge on the electron. Planck's constant $\hbar = h/2\pi$ is a basic unchangable constant of Nature with units of energy-seconds or energy per Hertz (i.e., each increment of cycles per second represents a tiny bit of energy $-$ the concentration or density of wavelengths per unit of time. A hyphen between units stands for multiplication).

UNITS: 1. 'SI' containing 'MKSA' Units ['System International' includes meters, kilograms, seconds, amperes]. This is the approved standard system for industry and students and everything else internationally (except sometimes in the U.S.A). In this system, electron charge is given as $e = 1.602 \times 10^{-19}$ Coulombs. The unit, Coulomb, is so large that it can represent a lightning bolt of electron charges. $\hbar = 1.054 \times 10^{-34}$ joule-seconds, $c = 3 \times 10^8$ m/s. And the $4\pi\epsilon_o$ is also required where ϵ_o is the "permittivity of free space" in units of $Coul^2/newton\, m^2$ (a first clue that a "Vacuum" might not be nothing– it has properties).

$$(2) \qquad \alpha = \frac{e^2}{4\pi\epsilon_o\hbar c} = \frac{(1.602\times10^{-19})^2}{4\pi[8.85\times10^{-12}][1.054\times10^{-34}][3\times10^{+8}]} = 0.0073 = \frac{1}{137}$$

(a pure number with no units remaining).

2. In particle physics, a favorite unit of energy is electron-volts, eV (charge times voltage difference). In this case "$e^2/4\pi\epsilon_o$ = 1.44 MeV fermi's (fermi = femtometer = $f = 10^{-15}m = 10^{-13}cm$), and $\hbar c$ = 197 MeV f . Dividing these numbers gives 1.44/197 =0.0073=1/137.

3. Some prefer 'ESU' (electro-static units [3]) where the charge on the electron is now $e = 4.803\times10^{-10}$ ('statCoulombs'), $\hbar = 1.0544\times10^{-27}$ erg-sec, and $c = 3\times10^{+10}$ cm/s. Then $e^2/\hbar c = 1/137$.

4. **Natural Units** use $c \equiv 1$ and $\hbar \equiv 1-$ no units. Nature doesn't care about our human scale (meters, joules, seconds...). Then the $\hbar c$ factor is just one. Particle physicists sometime prefer the rationalized 'Heavyside-Lorentz' units with these conventions and also that epsilon naught is unity as well ($\epsilon_o \equiv 1$, $\mu_o \equiv 1$ where $c \equiv 1/\sqrt{\epsilon_o\mu_o} = 1$)[1][2]. All that is left is then, $\alpha = e^2/4\pi$. Alternatively, the charge of the electron is $e = \sqrt{4\pi\alpha} \simeq 0.3$. So, charge e and $\sqrt{\alpha}$ are similar concepts; placing a $\sqrt{\alpha}$ at a Feynman vertex is similar to just placing an 'e' there. Since alpha has no units, charge e also has no units. It is also possible to choose units in which $4\pi\epsilon_o \equiv 1$ in which case $e \equiv \sqrt{\alpha}$.

5. There is a relationship between three decreasing basic reference sizes: the Bohr radius ($a_o = 0.53$ Angstroms $-$ about the radius of a Hydrogen atom), the Compton wavelength of the electron ($R_c = 386$ fermi's) and the classical electron radius $r_o = 2.82$ fermi's$-$ about 3 proton diameters. $\alpha a_o = R_c$ and $\alpha R_c = r_o$. In SI units, Bohr $a_o = [\hbar^2/m)(4\pi\epsilon_o/e^2)]$, $R_c = \hbar/mc$, $r_o = e^2/4\pi\epsilon_o mc^2$. So, take the 'big' size of an atom and divide by 137 to get the Compton size. Then divide again by 137 to get the 'classical' electron size. Although the electron is supposed to be a point particle with no size, it cannot be localized to within its Compton radius.

3. Using alpha

Feynman says in his book "QED" that the the probability for a charged particle to emit a photon is the square root of the fine structure constant. By this he means that the "vertex in the Feynman diagram" which has two external lines corresponding to a charged particle and one cross-joining wavey line corresponding to a cross-linking photon is proportional to $e \sim \sqrt{\alpha}$. But it takes two of these vertices to complete a virtual photon transaction, and that gives an amplitude proportional to alpha. Probability in quantum mechanics is amplitude squared. So the final result must contain an $\alpha^2 \propto e^4$. Indeed, when one looks at well known "cross sections" for scatterings, one does see equations $\propto \alpha^2$. Since alpha is a small number, alpha-squared is quite small. Using alpha this way implies that one is probably assuming natural units.

The electron energies in the hydrogen atom have well defined values calculated in every freshman physics class. But on top of these values are small changes called "spin-orbit fine structure" due to the interaction of the electron spin with its orbit around the nucleus. These values are proportional to alpha-squared [5]. This is one motivation for calling alpha a 'fine-structure constant.' "It appears in all first-order perturbation formulas for atomic energy levels."

Rutherford scattering is another freshman physics classical calculation yielding a probability of scattering an alpha particle (two proton charges) off of gold foil ($Z_{AU} = 79$ protons) at a certain angle θ as $\propto \alpha^2 Z_\alpha Z_{Au}/E^2 sin^4(\theta/2)$ [4]. This was the early experiment in England that led to the discovery of the atomic nucleus. A quantum mechanical calculation yields a similar result. Note that quarks scattering off a proton also follow a Rutherford type scattering formula, this time with a stronger strong-force coupling constant, ${\alpha_s}^2$. The Feynman diagrams are similar, but now a gluon goes inbetween rather than a photon.

Then there is Thomsom scattering of radiation off of free charges. Although the incident electromagnetic radiation has a wavelength and frequency, the power radiated at an angle theta is $\propto {E_o}^2 \alpha^2 sin^2\theta/(mc^2)^2$. In quantum mechanics, probability is amplitude squared. Similarly, in electricity, energy is electric field squared ${E_o}^2$. The total scattering cross section is the "classical electron radius" squared (where $r_o = e^2/mc^2 = 2.82$ fermi's)[3]. Books on particle physics multiply these results by $1 = \hbar^2/\hbar^2$ to get total cross section as $\propto \alpha^2 {R_c}^2$ where $R_c = \hbar/mc$ is the Compton wavelength of the electron.

The case of scattering off of bound charges (e.g., light off of the electrons in atoms in the atmosphere) is similar but with scattering $\propto r_o^2(f/f_o)^4$ where f is the frequency of the radiation and f_o is the resonant frequency of the oscillator. We are highly familiar with this '4th power of frequency' result from the enhanced scattering of blue over red light $-$ the reason that the daytime sky is blue.

Much of what we know about micro-nature comes from particle scattering experiments. The force causing deflections is due to Coulomb's inverse square law: $F = kQ_1Q_2/r^2$. This gives a force proportional to $e \times e = e^2$. A small particle moving at high speed towards a larger particle without any deflections would come closest by a certain distance called the 'impact parameter,' b. For charged particles, each angle of deflection, θ corresponds to some incoming energy and some value of b. Measurements look at angles of deflection, but they don't care about which direction $-$ they look collectively at deflections in all directions around a circle. The probability of deflection depends on a 'cross-section' circular area about the target particle: $\sigma = \pi b^2$. With the b^2 comes the factor$(e^2)^2$.
Feynman made up his interaction rules so they could agree with physics he already knew: how to get the e^4 final dependence.

4. Quantum Field Theory

In the subset of QFT called quantum electrodynamics (QED), there is a set of Feynman rules enabling an easier systematic approach to calculation. The rules relating to vertices include: attach a factor of $-ig = -i\sqrt{4\pi\alpha}$ to each joining point of three lines (two fermions and one boson). Calculations use complex numbers. It may be more appropriate to express that as $-ig\gamma_\mu$ where the gammas are the Dirac matrices which really represent math a little beyond just complex numbers into 'hypercomplex' numbers (like bi-quaternions). There is also a requirement that impose conservation of energy and momentum at each vertex. An internal photon is virtual, meaning that it can take on any momentum not necessarily adhering to classical conservation requirements. These force carrying virtual photons can borrow any amount of energy from the vacuum as long as they return it back to the vacuum fairly quickly.Then an integration is performed over all internal momenta. The net result of fairly complex calculations is an amplitude for an interaction. Then the Born rule of quantum mechanics requires that this amplitude be squared before it can represent any classical measurement result (like the angular scattering distribution over a great many particles). Each Feynman diagram corresponds to a unique integral. Only the lowest order in interaction involves the exchange of just one photon. More precision also allows the exchange of $n = 2$, 3, 4, or more photons together; but each case is suppressed by an amplitude $(e^2)^n$ or α^2 so that the easiest $n = 1$ case may suffice. However, the higher cases can cause integration to blow up to infinity. So a major accomplishment of QED was finding out how to interpret these infinities $-$ how to subtract then away from the underlying "correct" answer. This was called renormalization. Dirac never approved of it, but it results in great precision in agreement with careful measurements (such as the magnetic moment of the muon, for example). The case of $n = 1$ photon in a diagram doesn't mean that only one virtual photon is exchanged in an interaction. The calculation sums over all possible value of momenta of perhaps a great many photons.

A significant difference between QFT calculations and classical calculations is that individual scattering diagrams can be summed and interfere at the amplitude level. In the electron-electron (or Möller) scattering in Figure one for example, there is another associated figure to add. Since the electrons are identical particles, their identity could be swapped from the incoming to the outgoing pair. The two outgoing lines could cross over each other, and the observer wouldn't be able to tell the difference. So both of these diagrams have to be calculated and then summed and then squared for the final scattering probabilities (cross sections). Final formulae are more complex than for classical scatterings.

References

[1] Donald H. Perkins, Introduction to High Energy Physics, 4th Ed., 2000, Cambridge.
[2] Francis Halzen, Alan D. Martin, Quarks and Leptons, An introductory course in modern particle physics, 1984, John Wiley.
[3] John David Jackson, Classical Electrodynamics, 1962, John Wiley.
[4] Arthur Beiser, Concepts of Modern Physics, 1967, McGraw-Hill.

[5] Robert B. Leighton, Principles of Modern Physics, McGraw Hill, 1959, pg 199.
[6] Sacred Geometry, www.natures-word.com

5. Appendix:

On the separate issue of the golden ratio, $\phi = 1.618$ [6], it is does have great artistic and mathematical value. I do enjoy the Fibonacci series and the pentagram (etc....). It can be defined as the number whose reciprocal is obtained by subtracting one, i.e., $1/x = x - 1$, or $1 = x^2 - x$, or $x^2 - x - 1 = 0$, which can be solved by the quadratic formula:

$$ax^2 + bx + c = 0 = 1x^2 - 1x - 1, \; so\, x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{1 \pm \sqrt{5}}{2} = 1.618 = \phi \tag{3}$$

The special example of the compliment of golden ratio in a 360^o circle (yielding $\simeq 137^o$) depends on a somewhat arbitrary Babylonian convention of approximately 360 days in a year carried over to part of a circle. There can be many numerical coincidences, but only some are truly special.

ELECTRON SPIN AND SU(2)

DAVE PETERSON

ABSTRACT. In quantum mechanics, electron spin has many counter-intuitive features. Some of these are discussed loosely and using complex matrices. Lie Groups and Lie Algebras aid in representing the mathematics. The relevant group for quantum electron spin is $SU(2)$ which is based on hypercomplex quaternions ($i\times$ Pauli -'σ-matrices'). Some key concepts such as why electron 'spinors' have 4π symmetry remain intuitively opaque. The mass of the electron as well as most fermions and weak bosons is due to weak interactions with the Higgs Vacuum $-$ also describable using the SU(2) symmetry group. Definitions of key terms are provided at the end. [Preliminary].

1. INTRODUCTION:

To a new student, the discussion of particle spin in quantum mechanics (QM) presents much that appears strange. For example:

- No matter which x,y,z apparatus orientation is used to measure the deduced component of electron spin, the answer is always the same, $\pm\hbar/2$ units of spin angular momentum.
- If a z measurement of spin $s_z = +\hbar/2$ passes as a 'prepared state' into an x-oriented magnetic field, the deduced spin there will again be an even outcome of $+\hbar/2$ and $-\hbar/2$ units of spin now oriented in the x-direction.
- It takes two complete rotations of spin to return a spin wavefunction to its initial state (verified for neutrons in 1974).
- 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.
- In quantum field theory, the Dirac equation with its addition of relativity to QM magically derives antimatter and both the proper spin and magnetic moment of the electron even though the electron cannot be viewed as a 'spinning charge.' And also note that the expectation value of electron velocity is always the speed of light

Date: September 13, 2012, update to 10/20/12.
email: davepeterson137@gmail.com.

in any direction even if it is "just sitting there."

- The 'spin-statistics' theorem says that spin one-half requires the exclusion principle and electrons are fermions. Integer spin particles are bosons which can share the same space and quantum numbers.

My view for a long time has been that part of the strangeness of quantum mechanics is due to its wavefunctions living in "the square root of reality." This mainly comes from the basic "Born Rule" which says that the probability of a result is $P = \psi^*\psi$ so that a wavefunction, ψ, represents a strange concept called a "probability amplitude" which must be "squared" to represent classical reality. These square-roots (or "star-roots") introduce a basic importance of complex numbers and even hypercomplex numbers. We have a preconception that angular momentum requires something to rotate or orbit, but rotation transformations of vectors require 3×3 "rotation" matrices. The basic output of an electron spin measurement is experimentally noted to just have two states: spin-up or spin-down. We wish to represent this by a two component complex column vector called a "spinor." But, to transform this allows some rotation analog which only can be expressed as (complex) 2×2 matrices (involving $SU(2)$ Pauli spin matrices). A trivial $U(1)$ analogy is that the unitary rotation in a circle is represented by $\exp(i\theta)$, and the square-root of this is $\psi \sim \exp(i\theta/2)$. A full rotation of psi gives $\exp(i\pi) = -1$, and it takes two full rotations to get back to an initial $+1$. The "double covering" of the group $SU(2)$ is somewhat more mysterious.

One famous mathematician (Atiyah,[9]) 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 primary problem in understanding is that the words "real and "physical" are generally pre-empted to mean 'classically real.' If QM indeed uses complex and hypercomplex numbers in its mechanisms, then that would be "real" but also give misleading communication. I use the word "qreal" for **'quantum real.'** There is an amazing literature of mis-communication due to poor words (especially on the interpretation of QM). Popular books commonly mix them up because their goal is to show a strange world in normal terms.

Symmetry is a powerful concept in modern physics. Many laws of physics are invariant under transformations in spacetime or in "internal" spaces. In quantum mechanics, the mathematics of particle spin may be presented by matrix generators of the infinitesimal rotations of continuous rotation groups [the Lie Algebra of Lie Groups (Sophus Lie, 1873)]. The simplest example of a continuous smooth Lie group is the translation of the real number line $\mathcal{R}^1$ by the operation of addition by other real numbers '$+x$' which simply shifts the line to the left or right.

However, physicists often avoid elaborate rigorous mathematics and discuss rotations in terms of more familiar angular momentum matrices of quantum mechanics [1] (-including intrinsic spin, S, orbital angular momentum , L, and total angular momentum, $J = L+S$). They often may not mention Lie Groups and use "infinitesimal rotations" instead of Lie Algebras. Quantum mechanics promotes classical variables (abstract A) to operators ($\hat{A}$) [1] and discusses them using the resulting values of **'commutators'** $[\hat{A}, \hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}$. A tradition of frequently expressing quantum mechanics in terms of commutators '[ ]' was begun by Paul Dirac. Position and momentum do not commute $[x, \hat{p_x}] = i\hbar \neq 0$, and that leads to the uncertainty principle for those variables. [2] Complex variables are often used and are well accepted as being fundamental.

Angular momentum is defined as radius times linear momentum $\vec{L} = \vec{r} \times \vec{p} = (x, y, z) \times (p_x, p_y, p_z)$ by a cross product. So, for example, $L_z = xp_y - yp_x$ for motion about a z-axis. These relations as QM operators can give an inability to simultaneously measure say the x and y components of angular momentum together, $[L_x, L_y] = i\hbar L_z$. [3] Arguments such as this eventually lead to some of the strangeness surrounding electron spin. Although intrinsic spin is similar, it "does not depend on spatial coordinates in any way" [2]. The word 'wavefunction' is not appropriate, and only spinors or column vectors will be used.

2. Spin Operators:

A 'Stern-Gerlach' type experiment detects two states of electron spin in terms of their magnetic moments deflected by the gradient of an inhomogeneous magnetic field, ∇B. The detected spin angular momentum of these electrons have values of $s_z = +\hbar/2 \; and \; -\hbar/2$ which are usually labeled by 'up' versus 'down' or $|+\rangle$, $|-\rangle$ or $|\uparrow\rangle$, $|\downarrow\rangle$. The magnetic moment has value $\mu = -2\mu_B s/\hbar = \mp\mu_B$ or one 'Bohr-magneton' value. A basic concept of quantum mechanics is the idea of 'observables' which compatibly commute or don't commute as expressed by the commutator notation $[A, B] = AB - BA = 0 \; or \; \neq 0$. For example, In non-relativistic quantum mechanics (QM), electron spin is usually represented by spinor states, ψ, which are 2×1 column matrices and by 'Pauli operators' which are 2×2 matrices (Wolfgang Pauli, 1900-1958). These spin operators are designed to obey commutation relations:

$$[\hat{S}_x, \hat{S}_y] = i\hbar\hat{S}_z, \;\; [\hat{S}_y, \hat{S}_z] = i\hbar\hat{S}_x, \;\; [\hat{S}_z, \hat{S}_x] = i\hbar\hat{S}_y, \;\; [\hat{S^2}, \hat{S}_i] = 0, \tag{1}$$

[1] for example, $p_x \to \hat{p_x} = -i\hbar\partial()/\partial x$ and $\vec{p} \to \hat{p} = -i\hbar\nabla$.

[2] The preference for commutators is referred to by the 'Dirac correspondence principle for quantum operators.'

[3] I.e., $\vec{L} = \vec{r} \times \vec{p} \to -i\hbar\vec{r} \times \nabla$, form products like L_xL_y and regroup the xyz's and partial derivatives to get $(x\partial/\partial y - y\partial/\partial x) \propto L_z \propto \partial/\partial\varphi$ [11].

We can also define: $S^2 = {S_x}^2 + {S_y}^2 + {S_z}^2$, $\hat{S^2} = \left(\frac{\sqrt{3}\hbar}{2}\right)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Commutation relations like these apply to all the quantum angular momentum operators (e.g., S, L or J too). The basic Pauli matrices without the $\hbar/2's$ are labeled by the Greek letter sigma: $\hat{S}_i = \sigma_i \hbar/2$. Since the sigmas lack the division by 2, their commutation relation looks more like $[\sigma_j, \sigma_k] = 2i\sigma_\ell(\epsilon_{jk\ell}, cyclic)$. A common convention is to let the **spin-z operator be diagonal** with real integer eigenstate elements +1 and -1. That means that we let spin up and spin down in the z-direction to be conventional base states. The Pauli 'sigma' matrices are most often presented by:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

These Pauli sigma matrices are the **generators** of $SU(2)$; or more appropriately as $T_i = i\sigma_i/2$. These are related to the much older 1843 hypercomplex quaternions of Hamilton, $q_i = \pm i\sigma_i$ which form a continuous group Lie Algebra. We are free to use either Pauli matrices or quaternions, and physicists generally prefer the three σ_i's for parametrization of rotations represented in the group $SU(2)$ (see definitions at end). [4] The basis vectors or basis 'kets' for spin up and spin down are the column vectors:

(3)

$$|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ and\ |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.\quad Then\ e.g.,\ \hat{\sigma_z}|+\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

an eigenvalue form: $(Operator)(ket) = (value)(ket) = 1|+\rangle$. Once the diagonal convention is set for σ_z and σ^2, the other two forms are found simply by starting with something like $\sigma_x = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, and same for σ_y using b_{ij} 's, and substituting these into the commutators above to find equations whose solution gives the four values each for σ_x and then σ_y [5]. If the values in σ_x are chosen to be real, then the values of σ_y will be forced to be imaginary. Note that the square of any matrix is the identity or unit matrix, ${\sigma_i}^2 = 1$. And, $\sigma_x\sigma_y - \sigma_y\sigma_x = 2i\sigma_z$.

Characterizing electron spin in the 'y-direction' when bases are given in the z-direction can be done by finding the eigenvectors and eigenvalues of the spin operator, $\hat{S}_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ [2]. Note that In linear algebra [4], the eigenvalues ($\lambda's$ or "characteristic values" or "spectral values") of a linear operator or transformation matrix T can be found from a characteristic polynomial equation $\det(T - \lambda I) = 0$, where I is the identity matrix of diagonal ones.

So we must first find the roots λ of the characteristic equation $det|S_y - \lambda I| = 0$ which are the eigenvalues $\lambda_{1,2} = \pm\hbar/2$. So, the z-bases were found from $\pm\hbar/2$'s in the z direction, and now we also see $\pm\hbar/2$ in the y direction as well. The column eigenvectors are then

[4]{ The commutation relation for quaternions would look like: $[q_i, q_j] = 2q_k$ (cyclic) without imaginary i's.} The case of spin-1 would use "Gell-Mann" 3×3 matrices instead.

found just using algebra for

$S_y|+_y\rangle = \lambda_1|+_y\rangle$ or $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 1\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. The normalized result is simply $\alpha = 1/\sqrt(2)$ and $\beta = i/\sqrt{2}-$ as mentioned for the state in the introduction: $|\odot\rangle = a|\uparrow\rangle + ia|\downarrow\rangle$.

A vector in ordinary Euclidean or Cartesian coordinates $\vec{v} = (x_1, x_2, x_3)$ can be rotated about the origin in three dimensions so as to preserve its length $\ell^2 = {x_1}^2 + {x_2}^2 + {x_3}^2$. The relevant set of group rotations labeled as '$g's$' belong to the continuous 3-D rotation group O_3. These rotations can be described using consecutively applied "Euler Angles" ϕ_1, θ, ϕ_2, and g is a 3×3 matrix of sines and cosines of these angles. This may begin with a rotation ϕ_1 about the $z = x_3$ axis followed by a rotation of the z-axis through θ towards the new y axis (or about the new x axis). This group is homomorphic to another group $SU(2)$ specified by unitary matrices of order two and determinant unity [6]. Due to 'double covering' every rotation g of O_3 corresponds to 'two' matrices $+u, -u \in SU(2)$. An "infinitesimal generator" is formed by calculating the element by element 'slopes of matrices down to 0' from $g_r = [dg_r(\psi)/d\psi]_{\psi=0}$ evaluated at the group identity.
For example, let $g_z(\phi_1)$ be the first Euler rotation about the z axis by angle ϕ_1. Then,

$$g_3 = \left[\frac{dg_z(\phi_1)}{d\phi_1}\right]_o = \frac{d}{d\phi_1}\begin{pmatrix} \cos\phi_1 & \sin\phi_1 & 0 \\ -\sin\phi_1 & \cos\phi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\phi_1=o} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4}$$

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 in terms of cosines and sines by evaluating the <u>'exponential map'</u>:
$R_z = exp(\varphi g_3) \equiv I + g_3\varphi + (g_3\varphi)^2/2! + ...$ 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 (4)) happens to be the quaternion $q_y = i\sigma_y$. [5]

Repeating this procedure for the $u \in SU(2)$ gives three infinitesimal generators $g_i \, or \, u_i = \pm i\sigma_i/2 = \pm q_i/2$ ([6], p 9). Any element of the unitary matrices of SU(2) may be written as the exponential of a Hermitian matrix H, $U = e^{iH} \, where \, H = H^\dagger$ (see Definitions). In particular for us, $U = \exp(\theta^i i\sigma^i/2)$.

[5] So, ${q_y}^2 = -I$, minus the identity by definition, and ${q_y}^3 = -q_y$, ${q_y}^4 = +I$ repeat, etc. Use $sinx = x - x^3/3! + x^5/5!...$ and $cosx = 1 - x^2/2! + x^4/4!...$

3. 4π Spinor Rotations:

Why are **half-angles** used to describe rotations using 2×2 matrices like those based on Pauli matrices σ or quaternions q? (See Definitions). $R(\theta) = \cos(\theta/2) - \sin(\theta/2)(\vec{q} \cdot \hat{n})$. This observation also reflects the definition of a "spinor." We moved from rotations of 3-vectors to transformations of 2-component quantities, $\xi \to \xi' = R\xi$ where the ξ is a spinor which **reverses sign on a 360^o rotation**. It takes two full rotations to get back to the original + sign. We wish to consider a spinor as a 2×1 column matrix with electron spin-up in its upper part and spin-down in its lower part. The column size determines the matrix size, e.g., 2×2. There are many different approaches to this half-angle observation; but they all tend to be mathematical without very clear intuitive discussion:

Elementry 1-D Analogy: Intermediate students of mathematics should know that the second most important formula (after the Pythagorean Formula $a^2 + b^2 = c^2$) is $e^{i\pi} = -1$ and that in general $e^{i\varphi} = \cos\varphi + i\sin\varphi$. Suppose we take seriously the idea that in some sense, electron spin is like the square root of the mathematics for spinless 'particles.' Consider the elementary element of the circle rotations, $e^{i\varphi} \in U(1)$. What is the square root of this rotation through an angle φ?

$$\sqrt{e^{i\varphi}} = (e^{i\varphi})^{1/2} = e^{i(\varphi/2+(o,\pi))} = e^o e^{i\varphi/2} \ \& \ e^{i\pi}e^{i\varphi/2} = \pm e^{i\varphi/2} \tag{5}$$

The square root operation does two things: it takes us to half angles and it provides two solutions, a double cover for the mapping between the previous φ and the new $\varphi/2$. It takes a $720^o = 4\pi$ radian rotation to get to $+1$, i.e., $e^{i\varphi/2} = e^{i4\pi/2} = e^{2\pi i} = +1$.
The problem we really care about is 2×2 matrix operators of $SU(2)$. Is there any sense in which we might consider $SU(2) \sim \sqrt{SO(3)}$?

Quantum Angular Momentum: The "angular momentum" point of view [8] considers spinning electron rotation with $J = \hbar/2$ as a unitary rotation:

$$U_R(\phi) = exp\left(\frac{-i\vec{\phi} \cdot \vec{J}}{\hbar}\right) = e^{i\phi/2}. \tag{6}$$

So that, $U_R(2\pi) = e^{i\pi} = -1$, and $U_R(4\pi) = e^{i2\pi} = +1$. Then "half of the matrices that can be used to represent J or $U_R(\vec{\phi})$ are double-valued with respect to the angles vectors $\vec{\phi}$." So separate phi's that differ by a 2π rotation and can form a new group with twice as many elements as the rotation group $-$ a "covering" group. Each element in this group only has one element in the rotation group, but going the opposite way, there are two. The Pauli matrices generate a representation of the covering group. Due to 'double covering' every rotation g of O_3 corresponds to 'two' matrices $+u, -u \in SU(2)$.

Reflections: One book (Misner, Thorne, Wheeler) [12] describes this using rotations in terms of 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 that axis at the angle $\theta/2$."

Rotate the Apparatus: The Feynman Lecture Series [13] considers the 'filtering' of spinning objects with Stern-Gerlach apparatus in series with different orientation of trajectories and magnetic fields. He first discusses spin one (e.g., pions) and then spin one-half also leading to the $\phi/2$ angle amplitudes. The math is there, but the intuitive argument is not transparent. "For a rotation of 360^o about the z-axis, the amplitude to be in any state changes sign." His approach is based not on spin rotation but rather rotations and changes of orientations of the detection or perturbing apparatus [6].

Spheres: DeWitt [14] discusses this in terms of the interiors of two concentric spheres: "Because of the periodic dependence of (15) on ξ the parameter space of $SU(2)$ can be restricted to the inside of a 2-dimensional sphere of radius 2π, i.e., $\xi = \sqrt{(\xi^1)^2 + (\xi^2)^2 + (\xi^3)^2} \leq (2\pi)^2$. The center or origin is the identity element I; and outer radius is $r_2 = 2\pi$ and the whole boundary represents the single $SU(2)$ element $g = -I$. The parameter space of $SO(3)$ lies only in the inner sphere of radius $r_1 = \pi$. The parameter space of $SU(2)$ lives in an inner region $\xi \leq \pi$ and an outer region $\pi < \xi \leq 2\pi$. Each point ξ in the inner region corresponds to another point ξ' in the outer region so that a straight line connecting them passes through the origin and has length $\ell = 2\pi, \vec{\xi'} = \vec{\xi}(2\pi/|\xi|-1)$. $g_{SU(2)}(\xi') = -g_{SU(2)}(\xi)$. Altogether, then, the $SU(2)$ group manifold is $S^3 \subset E^4$. In a sense, this is double coverage of the inner sphere of $SO(3)$.

Classical: The appearance of half-angles in 2×2 matrix rotations has also been presented as a purely classical approach without reference to quantum mechanics [18]. One discussion is under "Cayley-Klein parameters" first used by Felix Klein for difficult gyroscope problems. He used a 2×2 complex unitary transformation matrix which turn out to depend on only three independent real values (the same number as for Euler angles for rotation of rigid bodies). The form for the infinite set of rotations, 'Q' $= Q(x_1, x_2, x_3)\, or\, Q(\phi, \theta, \psi)$ [7] , turns out to be the same as for the $su(2)$ group given by equation (12). Further mathematics shows that $Q(\phi = 2\pi) = -I$ rather than identity one. Also, an element of the 3-rotation $SO(3)$ corresponds to a **pair** of matrices $(Q, -Q)$ so that the Q matrix is a double-values function of the 3×3 orthogonal matrices.
The Q operators on 2-D complex spinors turn out to be more physically relevant to quantum mechanics. The use of half-angles and the double valued property go with half integral spin. Also QM requires self-adjoint or 'hermitian' operators, and $Q = Q^\dagger$ (complex conjugate of the matrix transposed) and leads to real eigenvalues.

Actually Measured: Experimentally, 4π spinor symmetry has been demonstrated by Rauch (1974) [15] using a perfect-crystal neutron interferometer. Mono-energetic neutrons can be

[6] His whole analysis seems horribly complex for lower level college classes (even for Cal Tech).

[7] Q for Quaternion. Cayley published his rotation work in 1843, and Klein (1849-1925) later obtained and applied similar results.

extracted from a nuclear reactor so that they have a wavelength near two angstroms and can be Bragg deflected (this is a neutron magnetic moment interaction with planes of nuclei also having magnetic moments). The interferometer can have three or more crystal blades where the first blade separates the incoming beam into two out beams which participate in a Mach-Zehnder pathway. [8] A great many experiments like this have subsequently been performed. Rauch used a magnetic field on one of the paths to alter the phase of the neutron to obtain interference when the paths join on the far side of the interferometer. A rotation of 2 pi yields a minus sign, and 4 pi gives a plus sign again. The path interaction is given by $H = -\vec{\mu} \cdot \vec{B} = -\mu\vec{\sigma} \cdot \vec{B}$.

4. **Definitions**

Physicists frequently avoid formal definitions of terms leaving them to be inferred rather than stated. Mathematicians are much better (but delete their motivations and scaffoldings). One may have to go through more than a dozen physics books to even get a glimmer of a clear definition. One reason for this is that **Nature** is the owner of definitions so that new discoveries and paradigms can change the definition. Math without interpretation is safer.

Intrinsic Spin: Spin is a quality of fundamental particles [9] which may be converted into classical angular momentum upon measurement [10] . Mathematically, spin is not a vector in R^3 nor a pseudo-vector; prior to interaction with an approximately classical object, it is "something else" (a "spinor"). Texts just say that spin represents an additional internal degree of freedom independent of spatial degrees of freedom, i.e., commutators with x, p, and L are all zero. Physics books often show pictures of angular momentum in terms of a Euclidean vector model for spin, orbital and total angular momentum, S, L, and J. This is useful and partly appropriate. S + L do combine like vector addition to give total angular momentum, J ["spin-orbit interaction"]. There is an analogy to gyroscopes and torque in a magnetic field leading to precessions.

But it fails to be a true picture for several reasons: 1) There is no spinning ball source for intrinsic spin; the outer portions of such a ball would have to be moving much faster than the speed of light. 2). Electron spin comes from the Dirac equation where relativity is essential, 3) Unlike vectors, the projection of S on a z axis is quantized, 4). One must rotate 720 degrees to return to the initial state. 5) In quantum mechanics, the use of complex numbers are a requirement in general and also for representing spin and spinors. 6) The electron appears to be a point particle with no size distribution for its charge.

[8] Modern teaching of QM now often begins with Mach-Zehnder interferometry, e.g., see [16].

[9] −− formerly "elementary" particles − before finding out that the proton and neutron were composite particles made of quarks and gluon fields.

[10] The Einstein-deHaas effect of 1906 indicated that a magnetic rod can be rotated by the application of a magnetic field so that spin is actually electron angular momentum. Photon AM has been directly shown to be able to rotate small mechanical particles.

Spinor: Roger Penrose [17] 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' [7]. 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 $|\alpha|^2 + |\beta|^2 = 1$.

Suppose electron spin is pointing (up) in some direction (θ, φ) with φ being an angle of rotation in the x,y plane and θ an angle from the z-axis. Then a unit vector in that direction is $\hat{n} = (x, y, z) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$. And $\vec{\sigma}$ is defined as $i\sigma_x + j\sigma_y + k\sigma_z$. Spin must be an eigenstate of $\hat{n}\cdot\vec{\sigma}$ with eigenvalue unity = 1, i.e.,

$$\begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 1 \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} e^{-i\varphi/2}\cos(\theta/2) \\ e^{i\varphi/2}\sin(\theta/2) \end{pmatrix} \tag{7}$$

The meaning of this according to [13] is that alpha and beta tell the amplitude for spin to be up along z and down along z. If n is aligned along z, then we have spin up z. If spin is aligned along x, then we have $1/\sqrt{2}$ amplitude for spin-up-z and also for spin-down-z. Note that under $2\pi = 360^o$ rotation for angle phi, $(\alpha, \beta) \to -(\alpha, \beta)$, so that it takes two full rotations to get back to the home state +1. Spin one would only require one full rotation, and spin 2 requires only half a rotation, π radians!.

Experimentally, 4π spinor symmetry has been demonstrated by Rauch (1974) [15] using a perfect-crystal neutron interferometer.

Lie Group:
A Lie Group (1873) is a "continuously connected group in which the parameter of the product of two elements are continuous, differentiable functions of the parameters of the elements..." [8]. The group is a differentiable manifold with group operations having smooth structure. Another statement is that a Lie group is an infinite group whose elements can be parameterized analytically. The product of two Lie groups is a Lie group, so the standard model $U(1) \times SU(2) \times SU(3)$ is a Lie group of dimension 12 or 1 photon + 3 vector bosons + 8 gluons.

Lie Algebra: A key focus of Lie Groups is to replace a global group with a linearized or local **"infinitesimal group"** which is now called the Lie Algebra [9]. The generators or 'tangent matrices' of the 'infinitesimal elements near the origin' of a Lie group form the basis of a Lie algebra [8]. The commutation relations between these generators "determine the main characteristics of the structure of the entire group, since in effect they specify how the group elements may be integrated to a finite distance from the identity element." The Lie Algebra is "closed in the sense that the commutator of any pair of generators is a linear combination of the generators." Infinitesimal groups exist because Lie groups are manifolds and hence have tangent spaces at each point [9]. A Lie Algebra 'g' has the same dimension as the Lie Group manifold $\mathcal{G}-$ they are locally isomorphic near the identity element. An **"exponential map"** gives a diffeomorphism between the identity neighborhood and a neighborhood of $\mathcal{G}$.

$$\exp(A) = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + ... \qquad or \lim_{n\to\infty}\left(I + \frac{A}{n}\right)^n = e^A. \tag{8}$$

e.g., $A = i\phi_1\sigma_z/2 \to U = e^A$. "The diffeomorphism established by the exponential map reduces the local study of a Lie group to that of its Lie algebra." Usually lower case letters are used for a Lie Algebra (often using German letters). E.g., for the rotation group $\mathcal{G} = SO(3)$ we use $so(3) = L(\mathcal{G})$. The Lie Algebra of $SU(N)$ is the set of all skew-Hermitian matrices with trace zero. In general, a product of two elements from the algebra to the group requires "the Campbell-Hausdorff formula" (not shown).
The simplest example of the exponential map may be for the $SO(2)$ group of 2-D rotations using all orthogonal 2×2 matrices with unit determinant. Its simple generator is mapped to: [14]

$$t = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \; g_{SO(2)}(\xi) = e^{\xi t} = \begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix}, \tag{9}$$

and again the generator is the quaternion, $t = q_y = i\sigma_y$.

Representation: Group representations describe groups using linear transformations of vector spaces in particular as matrices so that group operations are represented by familiar matrix multiplication. This reduces many group-theoretic problems to well understood linear algebra. So, given a group G with sample elements g, h, a representation of the group is the matrix form preserving the same multiplication rules as the original group it represents and is often labeled by $D()$. So $D(g,h) = D(g)D(h) \in D(G)$. "The mapping between G and D(G) is called a homomorphism. If the mapping is one-to-one, then it is called an isomorphism and D(G) is a faithful representation In case the mapping is not into a set of matrices but into some other algebraic structure, it is called a realization. In general a group can have many different representations" [14]. Each operator representation is formed by considering its action on a given set of basis vectors, $\{u_i\}$ so that $A_{ij} = \langle u_i|\hat{A}|u_k\rangle$. The special unitary Lie group $SU(2)$ is the set of all unitary 2×2 matrices with unit determinant.

Hamilton's Quaternions , H are also useful for representing the su(2) Lie algebra and obey ${q_1}^2 = {q_2}^2 = {q_3}^2 = q_1\, q_2\, q_3 = -1$ (1843). Another common labeling for the three hypercomplex numbers is i, j, k. These quaternions H have an older and different history predating and leading to the use of vector analysis. One matrix representation is:

$$q_o = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ q_1 = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix},\ q_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ q_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \tag{10}$$

In terms of the common **Pauli matrices** for $(\sigma_1, \sigma_2, \sigma_3)$ (2), we have $(q_1, q_2, q_3) = (i\sigma_3, -i\sigma_2, i\sigma_1)$. Another choice for $su(2)$ basis is $(-i\sigma_2, i\sigma_3, -i\sigma_1)$. Notice here that the sigma's (1,2,3) cycle backwards. If we wished (and preferably) we could chose as basis $q_i = i\sigma_i$ with all indices cycling forwards. In that case, the definition for the set of all members of the Lie Algebra $su(2)$ satisfies forms like:

$$su(2) = \left\{ \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix} : \alpha, \beta \in \mathcal{C},\ |\alpha|^2 + |\beta|^2 = 1 \right\} \qquad or \tag{11}$$

$$su(2) = \left\{ \begin{pmatrix} ix_3 & -x_2 + ix_1 \\ x_2 + ix_1 & -ix_3 \end{pmatrix} : x_j \in \mathcal{R} \right\} \qquad or \tag{12}$$

[10]. The bases are contained in this infinite set: e.g., $q_1 = q_x = i\sigma_x = i\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$. The $SU(2)$ sigmas and $SO(3)$ τ's have the same structure constants ϵ^{ijk} , and therefore $SO(3) \sim SU(2)$. the $so(3)$ and $su(2)$ are isomorphic Lie algebras, so $SO(3)$ and $SU(2)$ are locally isomorphic Lie groups (but not globally).

Physicists have a preference for the Pauli matrices instead of quaternions because (a) matrices have a natural interpretation as operators on vector spaces and b) the Pauli matrices generalize to higher-spin particles while the quaternions don't. These matrices are a basis of $su(2, C)$, which is the Lie algebra of $SU(2, C)$, which is isomorphic to $Spin(3)$, which is the double-covering group of $SO(3)$ [10], which is the group of rotations in 3-D space.

Orthogonal group in three dimensions, $O(3)$**:** The set of all 3×3 real orthonormal matrices with determinant +1. Rotations in general do not commute, so this group is not abelian. It is a continuously connected compact group. This is a generalization of $SO_2(R)$ diffeomorphic to the two dimensional circle: e.g.,

$$SO_2(R) = \left\{ \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} : \varphi \in \mathcal{R}/2\pi\mathcal{Z} \right\}. \tag{13}$$

Special Unitary Group of Dimension Two, $SU(2)$**:**

The group $SU(N)$ of degree N is represented by matrix multiplication of $N \times N$ unitary matrices with determinant one. Without the S, unitary matrices $U(N)$ have complex determinant values with modulus 1 and arbitrary phase [9]. "The group $SU(2)$ is isomorphic to the group of quaternions of norm 1 and is thus diffeomorphic to the 3-sphere." The infinitesimal generators often labeled T are represented by traceless hermitian matrices $tr(T_a) = 0\ and\ T_a = T_z^\dagger$. For example, the trace (sum of diagonal elements) of $q_3 = i\sigma_z$ is

zero.
Other Applications of $SU(2)$ in Particle Physics: 'Internal symmetries' like the $SU(N)$'s rotate fields and particles in an abstract isotopic space. One of the initial intentions was Heisenberg Isospin using SU(2) doublets. The Weinberg-Salam model used left-handed lepton doublets with interactions between leptons generated by intermediate vector bosons [7].

$$|\psi\rangle = \begin{pmatrix} p \\ n \end{pmatrix} \sim \begin{pmatrix} u^i \\ d^i \end{pmatrix}, \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \Phi_{Higgs} = \begin{pmatrix} \varphi^o \\ \varphi^- \end{pmatrix} \tag{14}$$

So, just as $SU(3)$ can be used for the Gell-Mann uds flavors as well as for color, the $SU(2)$ group can be sub-labeled for spin or isospin or "left" for weak interactions.

Angular Momentum Matrices:
For an infinitesimal length vector ϕ close to the identity element, an infinitesimal rotation is given by: $U_R \sim 1 - \phi \cdot Li/\hbar$ where L is a combination of the generators of infinitesimal rotations about the three coordinate axes through angles, ϕ_x, ϕ_y, ϕ_z. An example is unit spin angular momentum with 3×3 matrices S_x, S_y, S_z where e.g., $S_z = i\hbar g_3$ as shown in equation (4) above. [11]

Additional Details: The Jacobi identity for infinitesimal motions describes a basic rule for how three operators of a Lie Algebra commute together. Structure Constants, f, specify coefficients of commutation relations. E.g., for generators of a Lie Algebra of a group, $[J^a, J^b] = if^{abc}J^c$. For $SU(2)$, these are just the trivial permutation symbol, $f^{abc} = \epsilon^{abc} = (\pm 1, 0)$. But for larger groups like $SU(3)$, the f's are more numerous and complex. Specifying the structure constants enables reconstruction of the Lie Algebra. $SU(2)$ and $SO(3)$ have the same structure constants and are therefore locally equivalent.

The Dirac Matrices of Quantum Field Theory are built upon the Pauli sigma matrices, $\gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}$ and obey anti-commutation (+) $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$. Repeated multiplication of γ's gives a base of 16 "Γ" matrices such as $\gamma^5 = i\gamma^o\gamma^1\gamma^2\gamma^3$. The Γ's have infinitesimal generators called J^i for rotations and K^i's for 4-D relativistic boosts.

Lorentz Group: $O(3,1)$ not $O(4)$ due to sign change. The set of all 4×4 real matrices that preserve $s^2 = c^2t^2 - x^ix^i$. Including translations forms the Poincare group.

Hilbert Space: A Hilbert space or "state space" is a (possibly infinite, complex) linear set of square-integrable functions on configuration space possessing an inner product (e.g., $\langle\psi, \phi\rangle = \langle\phi, \psi\rangle^*, \langle\psi, \psi\rangle \geq 0$). QM 'wavefunctions' or states, ψ, are vectors in state space.
As a simplest example, a **qubit** or quantum-bit of information has a two dimensional

[11] Depending on the convention being used for Euler Angles which may alter signs.

state space given by $\{|0\rangle, |1\rangle\} = \{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\}$.

An operator or observable on Hilbert space may be 'represented' by a matrix, and the matrix operators must be Hermitian, $M = M_{ij} = M^\dagger \equiv \overline{M_{ji}} = M\ dagger-$ Hermitian conjugate or adjoint of M, interchanged rows and columns and complex conjugation. Hermitian operators have Real eigenvalues (a requirement for classical observation).
Note that the Pauli matrix $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is a Hermitian operator such that $\sigma_x|1\rangle = |0\rangle, \sigma_x|0\rangle = |1\rangle$. Any operator in a 2-D complex Hilbert space can be written in terms of Pauli matrices as $A = (1/2)(a_o I + \vec{a} \cdot \vec{\sigma})$, where $a_o = Trace(A), \vec{a} = Tr(A\vec{\sigma})$ [2]. In particular, $e^{i\theta\sigma_x} = I\cos\theta + i\sigma_x \sin\theta$.
Elements of the group $SU(2)$ may be written as [14]:

$$g_{SU(2)}(\xi) = exp(\xi^i i\sigma_i/2) = \cos(\frac{\xi I}{2}) + i\frac{sin\xi/2}{\xi}\xi^i i\sigma^i/2. \tag{15}$$

where $\xi = \sqrt{(\xi^1)^2 + (\xi^2)^2 + (\xi^3)^2}$. The g's form an infinite continuous set of 2×2) matrices.

Strong Isospin: The original idea of Isospin proposed for strong interactions in1932 by Heisenberg (and named in 1937 by Wigner) said that due to the charge independence of strong forces, the neutron and proton could be considered as two almost degenerate states of the same particle, the nucleon. Isospin can be considered as an analog to quantum spin and use the same mathematics in spite of not being any kind of spin itself. Like spin, there is a convention to talk about values of isospin in an artificial z or 3rd direction, I_3. Then the nucleon has isospin 1/2, with p and n having $I_3 = \pm 1/2$ and action by the Lie group $SU(2)_{spin}$. The modern explanation for nucleon isospin is in terms of quark content, $I_3 = (1/2)(n_u - n_d)$.
With higher experimental energy came new particles which could be grouped into multiplets still discussed using isotopic spin (isospin). For example, the Delta baryon quartet $(\Delta^{++}, \Delta^+, \Delta^o, \Delta^-)$ had max $I = 3/2$ decremented by ones to $I_3 = (3/2,\ 1/2,\ -1/2,\ -3/2)$ just like the old angular momentum decrements. The physical reasoning now is that quark transformation $u \to d$ lowers I_3 by one and lowers charge by one unit, (uuu, uud, udd, ddd). Strong isospin connects quark flavors globally without regard to handedness (chirality). For $Q = I_3 + Y/2$ where Y is 'hypercharge' or baryon number plus strangeness $= B + S$. $Y = 2(Q - I_3)$, $Y_u = 2(2/3 - 1/2) = 1/3$, $Y_d = 2(-1/3 + 1/2) = 1/3$, $Y_s = 2(-1/3 - 0) = -2/3$.

Weak Isospin: uses local gauge symmetry to connect the quark and lepton doublets of left-handed particles. To avoid confusion with strong isospin, the symbol T_3 is used, and all weak interactions preserve the value of T_3. The weak interaction cares about 'chirality' and works with left-handed fermion doublets, hence the group label $SU(2)_L$ (Glashow, 1961). There are no weak interactions for right-handed fermion singlets. The weak vector boson W^+ has $T_3 = +1$, $W^-\ has\ T_3 = -1$, $W^0\ has\ T_3 = 0$. $SU(2)_L \times U(1)_Y$ refers to weak isospin and weak hypercharge, $Y = a(Q - T_3),\ where\ Q = q/e$, and a is a free normalizer like 1 but sometimes 2 and "really" perhaps something else. For $a = 2$, $Q = T_3 + Y/2$. The electroweak Lie algebra generators are (T_1, T_2, T_3, Y_w). Y_w is a generator of the $u(1)$ Lie subalgebra, and T's are $SU(2)$ generators (e.g., quaternions/2). The correct charge depends on both T_3 and hypercharge or on both $SU(2)\, and\, U(1)\;\; as\; Q = T_3 + Y/2$ [12] . $T_3 = \pm 1/2$ for ν , and e_L has $Y = -1$, so $Q_e = -1/2 - 1/2 = -1$, and $Q_\nu = 1/2 - 1/2 = 0$. $Y = -2$ for right singlets [to give -1 for charge e, (Kaku p. 336)]. There is also a concept called "weak charge": $Q_w(neutron) \simeq -1$ which interacts with weak bosons. $Q_w(p) \simeq 0.048$, and $Q_w(e) \simeq 0.04$.(Find definition, how does this connect?).

Electron Mass: 'Elementary' particles are allowed to gain mass solely because of the weak force in which handedness or chirality becomes fundamentally important. Elementary particle masses appear to be generally larger than zero due to motion slower than light speed which in turn is due to a quickly repetitive "**zig-zag**ging space time motion. Each leg of the zig-zag is at light speed, but the net velocity is sub-light. This motion is similar to the old Schrödingers Zitterbewegung (relativistic jitter motion of 1931) but now depends on the Higgs field for the turnaround points and frequency. The physical electron, e, is presently viewed as alternating chiral left-handed electrons (zig L) [13] possessing weak charge and right-handed anti-positrons (R) which cannot interact weakly. A math example (using left and right chiral projections) is [22]:

$$\mathcal{L}(mass\, term) = -m_e\bar{e}e = -m_e\bar{e}[\tfrac{1}{2}(1-\gamma^5) + \tfrac{1}{2}(1+\gamma^5)]e = -m_e(\bar{e}_R e_L + \bar{e}_L e_R). \tag{16}$$

The electric charge (q = -1) and helicity are preserved during the motion; and the mass depends on the frequency of the zig-zagging. The Higgs field supplies net weak charge to the vacuum with a scale set by the distribution of fluctuating charges in the vacuum. For massive W bosons, the Higgs gauge boson and the massless weak gauge boson flip back and forth into each other (e.g., ϕ^+ and W^+) . Penrose says that the Dirac equation for 4-spinors can be written as an equation coupling two Weyl 2-spinors, each acting as a kind of source for the other with a Higgs field coupling constant describing the strength of the interaction between the two. Zig is a source for zag, and zag is a source for zig. L-Fermions have weak isospin charge $T_3 = +1/2$ for the uct quarks and neutrinos, ν, while $T_3 = -1/2$ for the dsb quarks and electron-lepton. The massive boson W^+ has $T_3 = +1$ and W^- has -1.

Mass and Higgs Field: The principle of Lagrangian Least Action is a highly fundamental concept of Modern Physics. In relativistic quantum mechanics, a prime example is the Klein-Gordon equation and its Lagrangian:

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \tfrac{1}{2}m^2\phi^2 \Rightarrow \qquad \partial_\mu\partial^\mu\phi + m^2\phi = 0. \tag{17}$$

A term with field-squared has its coefficient interpreted as particle mass ($m^2/2$). A Higgs field, h, is a local variation about a new displaced vacuum ground state field, φ_o resulting

[12] This is more appropriately expressed as currents for electromagnetic currents, weak isospin currents, and weak hypercharge currents, $j_\mu{}^{em} = J_\mu{}^3 + (1/2)j_\mu{}^Y$

[13] Leptons form left-handed weak-isospin doublets, $L = \begin{pmatrix} \nu \\ e \end{pmatrix}_L$. Left handed states are projected from physical states by: $e_L = \frac{1}{2}(1-\gamma^5)e$ and $\nu_L = \frac{1}{2}(1-\gamma^5)\nu$.

from symmetry breaking of a previous vacuum field (a "Mexican Hat" [14] or "Wine bottle bottom" shaped potential field profile is usually assumed, and the field moves from the zero-center to a lower valley ground state, v, reference location). The mathematics yields a Lagrangian term $\mathcal{L}_{term} = -\lambda v^2 h^2$ so that a new mass is generated as $m_{higgs} = |v|\sqrt{2\lambda}$ from the coefficient of h^2 perturbation field from $\varphi(x) = v + h(x)$. The mass is proportional to the value of the new field, v. Similar calculations show that the mass of the W's and electron are also proportional to the new vacuum ground state field, $\varphi_o = v$ but now as coefficients of $|W|^2$ or $|e|^2$ fields squared.

References

[1] Harry J. Lipkin, Lie Groups for Pedestrians, North-Holland Publishing Co., Amsterdam, 1965.
[2] David McMahon, "Quantum Mechanics DeMystified," 2009, McGraw-Hill.
[3] David McMahon, "Quantum Field Theory DeMystified," 2008, McGraw-Hill.
[4] Kenneth Hoffman, Ray Kunze, Linear Algebra, 2nd Ed., 1971, Prentice-Hall, Inc. p. 183.
[5] "The Magnificent Matrix Mathematics of Spin," Chapter 7, 7 Aug 2010, http://eg.bucknell.edu/ ∼ koutslts/Ph332/Text/Chap7_JQPMaster.pdf
[6] Moshe Carmeli, Group Theory and General Relativity, McGraw-Hill, 1977.
[7] Michio Kaku, Quantum Field Theory, Oxford Press, 1993.
[8] Leonard I. Schiff, Quantum Mechanics, Third Ed., McGraw-Hill, 1949, 1955, 1968.
[9] Lie Group, Wikipedia http://en.wikipedia.org/wiki/Lie_group. 9/20/12.
[10] Willard Miller, Jr., Symmetry Groups and their Applications, Academic Press, 1972.
[11] Louis Harris, Arthur Loeb, Introduction to Wave Mechanics, McGraw-Hill, 1963.
[12] C. Misner, K. Thorne, J. Wheeler, GRAVITATION, W.H. Freeman and Co., 1970, Chapter 41.
[13] Feynman, Leighton, Sands, The Feynman Lectures on Physics, Vol III, Quantum Mechanics, Chapter 5 (Spin One), Chapter 6 (Spin One-Half), Addison-Wesley, 1964. Quote pg. 6-9.
[14] B. DeWitt and J. Smith, Field Theory in Particle Physics, Volume 1, North-Holland, Elsevier, 1986. pgs. 426-432.
[15] Rauch H., Treimer W., Bonse U., Phys. Lett. A47, 369 (1974). Also see Neutron Interferometry, Oxford, 2000.
[16] Benjamin Schumacher, Quantum Mechanics: The Physics of the Microscopic World, The Great Courses.
[17] Roger Penrose, The Road to Reality, Alfred A. Knopf, N.Y., 2005
[18] Herbert Goldstein, Classical Mechanics, Addison-Wesley, 1950.
[19] Theodore Frankel, The Geometry of Physics, An Introduction, Second Edition, Cambridge Press, 1997, 2004.
[20] Sadri Hassani, Mathematical Physics, A modern introduction to its foundations, Springer, 1999.
[21] Morton Hamermesh, Group Theory and its application to physical problems, Addison-Wesley, 1962.
[22] Francis Halzen, Alan D. Martin, Quarks & Leptons: An introductory course in modern particle physics, John Wiley, 1984.
[23] A. Zee, Quantum Field Theory in a Nutshell, Princeton, 2003.

5. Additional Mathematics:

The group $SU(2)$ is simply connected, but $SO(3)$ is not simply connected [19]. Topologically, $SU(2)$ is the 3-Sphere, S^3, which is the set of unit vectors in $C^2 \approx R^4$, $S^3 =$

[14] $V(\phi) = \mu^2\phi^2/2 + \lambda\phi^4/4 \; for \; \mu^2 < 0, \; and \; \lambda > 0.$

$\{(z_1, z_2)^T : |z_1|^2 + |z_2|^2 = 1\}$, i.e., $x_1^2 + y_1^2 + x_2^2 + y_2^2 = 1$. $SU(2) \approx S^3$. $SU(2)$ is a fiber bundle over $SO(3)$ with fiber always consisting of exactly two points $\pm u$. That is, $SO(3)$ is the projective space RP^3 which results from S^3 by identifying pairs of antipodal points. When $SU(2)$ is thought of as the double cover of $SO(3)$, it is called the spinor group $Spin(3)$. If Lie group parameters vary over a finite range (closed), then the set of elements of the Lie group (the group manifold) is said to be compact [21]. "A continuous function on a compact set is bounded." "The unitary $U(n)$, orthogonal $O(n)$, special unitary $SU(n)$, and the special orthogonal $SO(n)$ groups are all compact," and "Compact Lie group can be represented by matrices" [20].

6. Broken Symmetry Higgs Comments:

Textbook calculations of the mass generations from the Higgs field usually start with:
1) a simple single real field, φ, having a quartic or "Mexical Hat" profile,
2) a simple complex field $\varphi = \varphi_1 + i\varphi_2$,
3) spontaneous breaking of a global single complex gauge symmetry,
4) finally the actually appropriate case of the breaking of a <u>Vacuum Doublet field</u> (for the electroweak SU(2) case) [15].
Glashow in 1961 first proposed the joint group structure $SU(2)_L \times U(1)_Y$ allowing the possible existence of the photon and neutral current along with weak interactions (four generators for four fields). An SU(2) doublet of complex scalar fields covers 4 scalar particles and is invariant under global SU(2) phase transformations [22]. This is broken ($\rightarrow$)into just one real field (v+h) and then to its fluctuations about the ground state.

$$\phi = \begin{pmatrix} \phi_\alpha \\ \phi_\beta \end{pmatrix} = \sqrt{\tfrac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \rightarrow \sqrt{\tfrac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \approx \sqrt{\tfrac{1}{2}} \begin{pmatrix} H_2 + iH_1 \\ v + h - iH_3 \end{pmatrix} \tag{18}$$

The appearance of the local fluctuations (H's) comes from the application of the SU(2) transformation $\exp(i\tilde{\tau} \cdot \tilde{\mathbf{H}}(x)/v)$ to the real $\phi(x)$ (Weinberg, 1967). The three massless-Goldstone Higgs gauge fields (H_1, H_2, H_3) combine with (get 'eaten by') [16] the massless W 'Goldstone' boson fields (W^1, W^2, W^3) to produce massive weak physical bosons (mass eigenstates) and a massive scalar h (the Higgs Boson). The H fields are then re-zeroed leaving just the v+h field. Notice that the word "Higgs" is used in a variety of ways, none of which really makes Higgs happy: a quartet of displaced vacuum fields, the background ground field, a perturbing field, and the quanta (massive scalar boson) of the h-field.

[15] It is possible that there are more than one Higgs doublets. At this point it doesn't seem to be required, but it is also not yet disallowed.

[16] The Higgs mechanism of a massless gauge field becoming massive by eating a Nambu-Goldstone boson was first used in condensed matter physics where it is known as the Anderson mechanism [23].

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 $e^{-r/a_o} \rightarrow e^{-\alpha r/a_o} \simeq e^{-1.238r/a_o}$. [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 $E_h = 27.21\,eV = 2.626MJ/mol$ 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 cell 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 $< KE >=< p^2 > /2m = +\hbar^2/2m{a_o}^2 \simeq +13.6eV$. But the expectation value of potential $< V >=< -e^2/4\pi\epsilon_o r >=< -\hbar^2/a_o m r >= -\hbar^2/{a_o}^2 m \simeq -27.4$ 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 $\hbar \to 0$).

Discussion 1 (June 21, 2011 - 7/7/11, DP, PRELIM version 2).

"Test of Quantum Entanglement" – Aspect Experiment:

Importance: Quantum Mechanics is strongly counter-intuitive. The Aspect Experiment of 1982 was one of the first tests to show that quantum reality is non-local (appears to involve "faster than light" communication between entangled particles) and also demonstrated "delayed choice" (detector settings being set after particle emission and during flight). It really confronted us to ask, *"HOW can Nature be like that!"* In these tests, the polarization of each photon in a pair is not determined until a detector sees it and causes an "instantaneous action at a distance with regard to the measurement of the polarization of the other member of a given pair" ("spooky action at a distance"). Entanglement and superposition are key to the new fields of quantum computation and quantum cryptography.

Entanglement Definition: Quantum entanglement or "non-local connection" of two or more objects refers to the strange non-classical "linking" of the objects of a system so that one cannot adequately describe the quantum state of a constituent without full mention of its counterparts, even if the individual objects are spatially separated. This interconnection leads to correlations between observable physical properties of remote systems, often referred to as nonlocal correlations. "Einstein famously derided entanglement as ``spukhafte Fernwirkung" or "spooky action at a distance". Schrodinger's term was 'Verschrankung' or 'cross-linking' (or `shared enclosure'). John Cramer treats entanglement as joint communications back and forth in time between particles and their source (e.g., Fig. 6 below). Photons know about their future joint detections because "they have already been there!" Even in the de Broglie/Bohm "pilot wave" case, it would seem that some back-and-forth communication is necessary merely to establish the relevant wave-function prior to using it to create a quantum potential and establish a relevant final pathway.

Most modern tests of quantum entanglement mainly use light photons and measure their polarization (e.g., up-down, or sideways). In my drawer at home, I've always had a clear calcite crystal and also a little package containing three cheap green plastic Polaroid squares (a "3P" demonstration). If two Polaroids are aligned, they transmit light; but if they are turned against each other (crossed-polarization), not much light gets through and the squares are fairly dark. If two Polaroids are crossed and the third one is added on the outside, the result is a little darker yet. What happens if the third Polaroid square is placed between the other two but at a diagonal 45° angle? It gets lighter!—more light is transmitted! This defies intuition. The electric amplitude of the light from the first Polaroid gets projected onto a 45 degree line to 70% of its value and then gets <u>projected</u> (i.e., cosine) again to the last crossed Polaroid to get another 70% value [cosine(45°) ~ 0.7 [15]] . For just two Polaroids at an angle θ to each other, the transmitted intensity is amplitude squared or $\mathcal{I} = (\cos\theta)^2 = \cos^2\theta$ (Malus' Law, ~ 1810). For aligned squares, $\cos^2 0° = 1$ = 100%, and for crossed Polaroids $\cos^2(90°) = 0$ (no transmission). For two sets of 45°, the output intensity is then about 50% of 50% or near 25% transmitted. This also applies for a collection of single photons one at a time, except that what gets measured now are total quantum counts rather than variable intensity. So quantum mechanics of polarization correlation expects a Malus' Law result. The Frenchman Malus also discovered that light can be polarized by reflection from a tilted glass surface. For the new entangled photon case, two correlated photons moving in opposite directions through two separately oriented Polaroids act like one photon moving through two polaroids with different tilts.

And the calcite—well, it is "birefringent" so that different polarizations of light get bent by different amounts. A crystal on top of a printed word shows two staggered words, and a crystal on top of a printed dot shows two dots because the light from the print contains all directions of polarization at once. So all polarizations from a page of paper get altered to just two: one parallel to the crystal axis and one perpendicular (called ordinary and extra-ordinary)-- and these get bent by different amounts to give two dots. Rotate the crystal, and one dot will rotate about the other (neat!). A Polaroid on top of the dot and the crystal on top of that at just the right angle will only show one dot. Calcite can be used to detect and measure polarization.

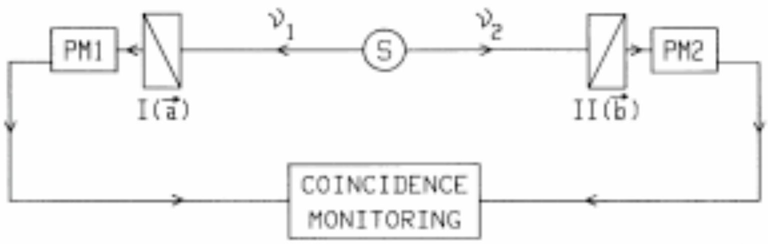


FIG. 1. Optical version of the Einstein–Podolsky–Rosen–Bohm *Gedankenexperiment*. The pair of photons ν_1 and ν_2 is analyzed by linear polarizers I and II (in orientations $\vec{a}$ and $\vec{b}$) and photomultipliers. The coincidence rate is monitored.

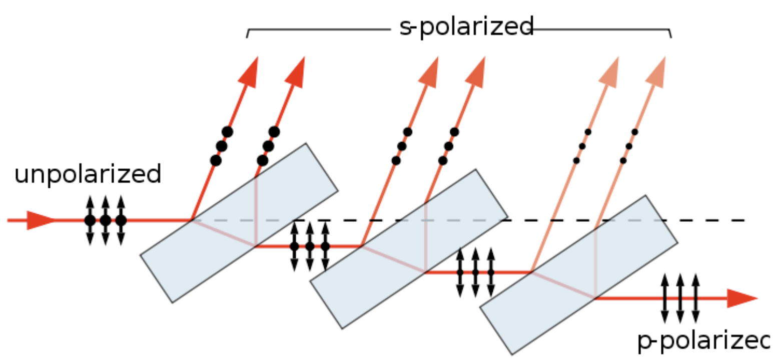
Figure 1 shows an idealized 'single channel' EPR test [4] modified from original Einstein $\Delta x \Delta p$ test to Bohm spins and now to the modern case of photon polarizations. A little more detail is supplied from the particular test by grad student Stuart Freedman and post doc John Clauser ["**FC**," 1972, 3] in Figure 3 below. This was the first application that took Bell seriously and was a partial test of Bell's ideas.

John Bell opened the window to testing quantum mechanical entanglements—but his thinking is often considered to be difficult. By assuming with Einstein that quantum mechanics ("QM") is really Local with no superluminal communications, he derived formulas that conflict with actual measurements thereby showing that QM must be non-local so that communications do not diminish with distance and act instantaneously. A few popular writers like Nick Herbert [12] offer more elementary views like Bell based on the assumption of locality. In most "Bell tests," long distance communications seem to be instantaneous and appear to violate relativity.

Herbert says, "Bell's theorem is easier to prove than the Pythagorean theorem taught in every high school." Like "Freedman and Clauser" FC-1972 and an improved Aspect 1981 version of FC, he considers visible 'blue' photons (for mercury, but violet for calcium) and light green photons emitted at the center of an apparatus and moving to the left and right through polarizers into detectors—a green detector and a blue detector ("B" and "G," -- appropriate because color filters were actually used in the Bell type experiments near the detectors). If the B and G polarizers are aligned together, then ~100% photon detection correlation can occur. A "B" photon in the B detector guarantees a G photon in the G detector. He considers calcite detectors, but photomultipliers and polarizers work too. The polarization directions can be set separately from up-alignment with angles ϕ_G and ϕ_B. Quantum mechanics ("QM") says that the output polarization correlation PC counts will only depend on the difference angle $\Delta\phi$ and not on each local setting separately. This implies that real QM detection near the polarizers know about each other's orientations so they can behave jointly. The correlation results for QM is a cosine-squared curve versus a triangular plot for local hidden variables. They give the same results at 0°, 90° and sometimes 45° -- but the QM results are enhanced in-between. So asking what happens at 22.5° or 30° would be revealing. $\text{Cos}^2(30°)$ = 75% correlation of photon detections or 25% misses. The Bell/Einstein locality supposition is that turning the Blue polarizer can only change the Blue message and not the green – a reasonable assumption, but wrong. If the blue polarizer is turned +30° and the green polarizer is turned -30° then locality would predict a 25% + 25% = 50% misses and 50% correlation. BUT, reality is $\cos^2(2\text{x}30°)$ = 25% net correlation or 75% misses. A correction to locality is that if blue is an error and green also happens to be in error, then that is a correlation again—a correction factor to the estimate. The result is then a net error rate of 50% OR LESS—an inequality, which is an example of a "Bell Inequality." So getting 75% misses would violate the inequality and support quantum mechanics. Freedman and Clauser got 75% misses and contradicted locality. The Bell inequality is an indirect statistical measure of locality that assumes that reality is reasonable.

Bell experiments are often described as tests of Bell Inequalities. These are often shown as complex "Venn diagrams" shown on paper for overlaps of different 'sets' of events (settings A or A', settings B or B', or "other" (e.g., no polarizer present). Sometimes these are elaborately colored to keep tracks of all the overlaps [14]. The locality assumption is that all set partitions are logically independent. A setting might be A = Bell test angle of 0, 22.5°, 45°, 67.5° or 90°. Counts in each setting category are measured, and an inequality involving set-combinations is stated as a test metric.

There are many different examples of Bell inequalities (e.g., the 'CHSH' inequality for Clauser's test). One of the simplest is, "the number of objects which have parameter A but not parameter B plus the number of objects which have parameter B but not parameter C is greater to or equal to the a count of number of objects which have parameter A but not parameter C," i.e., $N(A^+B^-) + N(B^+C^-) \geq N(A^+C^-)$. These statements are derivable using simple logic (not shown here).
An example of the categories could be A = male, B = tall, C = blue eyes – they could be any parameters. Bell was thinking mainly of particle spins for Bohm's version of EPR. For photons, A could be "polarization up" and detector up or ϕ= 0°, B photon up but detector at ϕ = 22.5°, C photon up but detector at ϕ = 45°. If an electron spin experiment discusses a test using an angle θ, a polarization test would use half that angle. There is an assumption that electrons have a spin in a given direction even if we do not measure it (but this is not true).



<u>Figure 2</u>: Stack of Plates polarizer with each plate tilted at "Brewster's Angle". Since the amount reflected at each glass is small, it may take 20 plates to obtain much output polarized light. [WIK]. The FC apparatus using this was a huge kludge on sawhorses. "p" means transmitted light in the **p**lane of incidence, and "s" is $\perp$.

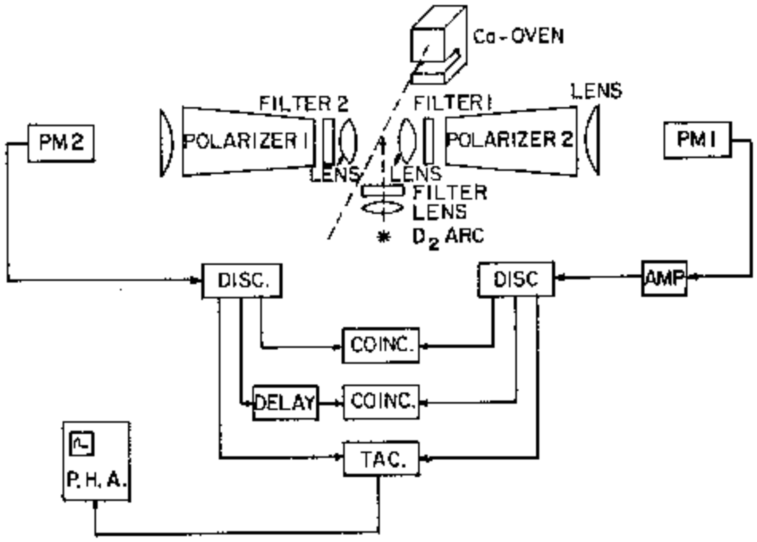


Figure 3: Freedman-Clauser (FC) test using excited state calcium-40 atoms decaying to a ground state which produces two correlated photons moving to the left and right through two differently tilted polarizers and then to photo-multipliers PM. [From reference 3]. The large polarizers are stacks of tilted glass plates. (Figure 2 above).
Then Figure 5 shows the more complicated Aspect Test [4].

Correlations between the polarizations of pairs of photons that are created in an atomic transition were studied by Clauser and Stuart Freedman in 1972 at the University of California at Berkeley. They performed measurements on the correlations and showed that Bell's inequality was violated [1] thus showing that photon pairs were entangled. Because this was a first and difficult experiment, they had several "loopholes" in this experiment such as not having a fair sample of all photons emitted by the source (the detection loophole) or possibility of un-noted causal connections (the locality loophole).

The1982 French 'Aspect experiment' improved on 'Clauser and Freedman's experiment by using a two-channel detection scheme to avoid making assumptions about photons that were detected. They also varied the orientation of the polarizing filters during their measurements – and in both cases Bell's inequality was violated [9]." "The locality loophole was closed in 1998 by Zeilinger and colleagues at the University of Innsbruck, who used two fully independent quantum random-number generators to set the directions of the photon measurements."

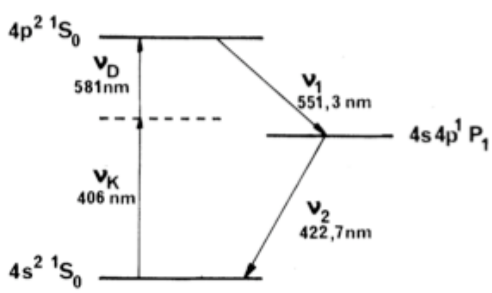


Figure 4: Excited energy state diagram for the **Two photon cascade:** Pump an Atomic beam of calcium atoms up to an electron excited state which then emits a cascade green and (almost

immediately) a violet photon with correlated polarizations. This was used by Aspect and by Freedman & Clauser [who in turn borrowed the idea and apparatus from Kocher and Commins] but with two-photon absorption from a Kr and Dye laser. The polarizers observing these photons are 'pile of plates' [Aspect 1981 PRL 47 p460]. {The names of the states above refer to principle and orbital quantum numbers and their occupation by electrons. Going from net spin zero to zero requires two photons carrying away one unit of angular momentum each}.

Figure 5: Aspect's 'delayed choice' EPR test modified from Figure 1 above. This picture is mainly complicated by showing the quick switching devices for the left and right photon paths . (Fig 2 in Aspect's Article).

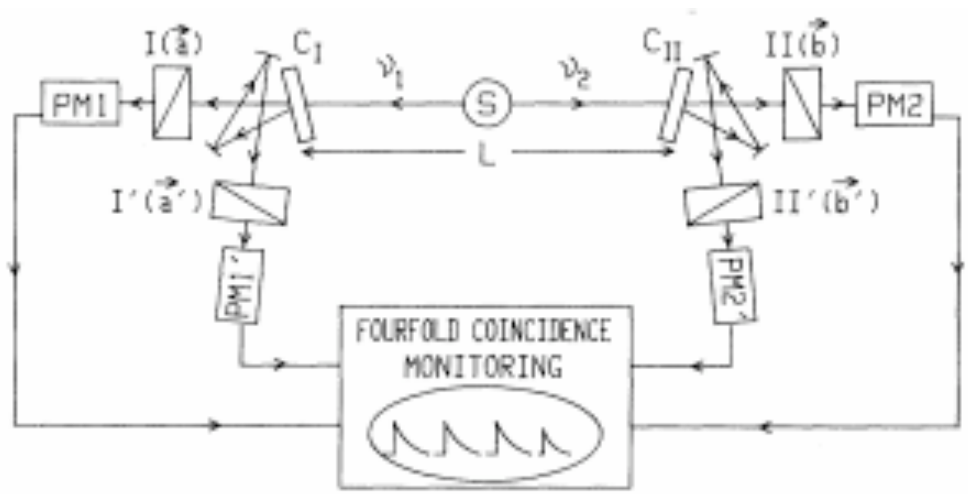


FIG. 2. Timing experiment with optical switches. Each switching device (C_I, C_{II}) is followed by two polarizers in two different orientations. Each combination is equivalent to a polarizer switched fast between two orientations.

Relevant History:
1935: Einstein, Podolsky, Rosen Classic paper "EPR". This represented Einstein's 'local and real' thinking about entanglement, but he didn't know about the paper being submitted in his name. He believed the Copenhagen interpretation of quantum mechanics to be incomplete. That is, QM differs from the 'obviously true' local realism and hence must be incomplete and have unrecognized 'hidden' variables. It was believed that Einstein lost his debates to Niels Bohr so that EPR ideas were not generally accepted at that time, and world-wide quantum introspection ceased for decades.

1964: John Bell's theoretical papers showing that ideas on entanglement and locality could actually be measured! This is sometimes called "the most profound discovery of science." He revived de Broglie/Bohm views versus the Copenhagen interpretation. Note that Bohm theory has nonlocal hidden variables (position and velocity) and is viable. Bell's theorem or Bell's inequality "is a no-go theorem, loosely stating that no physical theory of local hidden variables can reproduce all of the predictions of quantum mechanics." It is important to note that Bell's statements involve logic only and have nothing directly to do with quantum mechanics. The failure of Bell inequalities means that reality doesn't follow what we might call conventional assumptions and logic.

1972: Freedman and Clauser ("FC") succeeded for the first time in preparing two particles that exhibited the strange condition predicted by quantum theory called 'entanglement'. This FC test was improved in a 1981 Aspect test (without delayed choice). [They used a photon cascade from calcium but later experiments use 'down conversion'].

1978: John Wheeler's version of a delayed choice thought experiment in which the method of detection can be changed in flight after the photon passes the double slit and force the photon to decide if it is a particle with path or a wave with interference (verified by Aspect in 2007).

1982: **Aspect's** test based on Freedman and Clauser's test but with better precision, fewer loopholes, and delayed choice of detector polarization orientation. This test finally forced the world to take entanglement seriously and started a rebirth of quantum debating.

Aspect Test:

Alain Aspect was an early convert to the possibilities of John Bell's theorem. He wanted to show that non-locality was indeed an essential element of quantum mechanics and that Einstein's belief in locality was misplaced. His first goal was to replicate previous tests for one and two channels but with more accuracy and fewer loopholes. And then he wished to perform the Wheeler random delayed choice test evolved further by Bohm, Aharonov, and Bell. Testing quantum mechanics was academically-politically risky, so he got a blessing from Bell, "You must be a very courageous graduate student" [13].

In Paris, Aspect used a long distance between detection stations of about 12 meters allowing for a long transit time for light of 40 nanoseconds. There are rapid switches near each detection station that decide one of two alternative measurements to make. They are very clever 'electro-optical-acoustical' transducers, driven in phase, creating ultrasonic standing waves in a slab of water through which the relevant photon must pass using a frequency of about 25 MHz (the frequency is different for the two stations). The periodic density variation in the wave acts as a diffraction grating: If a photon is pictured as a localized 'wave packet' (length in time ~5 nano-seconds) that arrives at say station-1 when the wave has a node in-between its peaks, it is transmitted straight through the slab and enters a polarizer set in direction 'a'. If on the other hand it arrives at an antinode (periodic density peaks of counts or probability), then it undergoes Bragg diffraction and is directed into a polarizer set at a' (as sketched in Figure 5 above). Light quantum Photons incident at intermediate phases of the wave are deflected into neither polarizer and are thus missed in the counting. This experimental idea is just amazing to me. The period of switching between the alternative choices (a quarter period of the transducers) is about 10 nsec., short compared to the transit time of light between the stations. To the extent, then, that one can regard the switching as a "random" process, the locality loophole is blocked. The data obtained in ref. [4] violate the Bell predictions by many standard deviations. Truly random choices were conducted much later.

How is it possible for the two polarizers to 'know about each other' and know about recent changes in testing instantaneously? This appears not only to violate common sense but also relativity limited by the speed of light, c . One possible explanation was proposed by John Cramer in 1986 [6,7,8]. He has a view of quantum mechanics called the "Transactional Interpretation." The math of quantum mechanics is straightforward, but there are many different interpretations of what the math represents in the physical world. Cramer essentially says that a photon is it own antiparticle and can travel backwards in time as well as forwards. Feynman also considers any antiparticle as a particle **traveling backwards in time (fig 6)**. When an emitter wants to emit a photon, it first sends an offer wave forwards in time. An absorber then sends a confirmation wave backwards in time to the emitter. The communication process continues until a completed transaction occurs and the proper quantum numbers are delivered to the absorber. The emitter knows what kinds of tests it is going to encounter in the future because its wave function 'has already been there.' In quantum mechanics, the offer wave is called 'psi' and the confirmation wave is 'psi-star' (ψ, ψ^*-- complex conjugation represents time reversal). The probability of the transaction is then very naturally the product P = $\psi^*\psi = |\psi|^2$. This forward-backward propagation also represents normal electromagnetic wave propagation as 'half-advanced half-retarded' electric fields going forward and backward in time (R and A waves—the Wheeler-Feynman electrodynamics).

The following space-time diagram shows how two well-spaced detections can have instantaneous quantum communication using forward-backward communications.

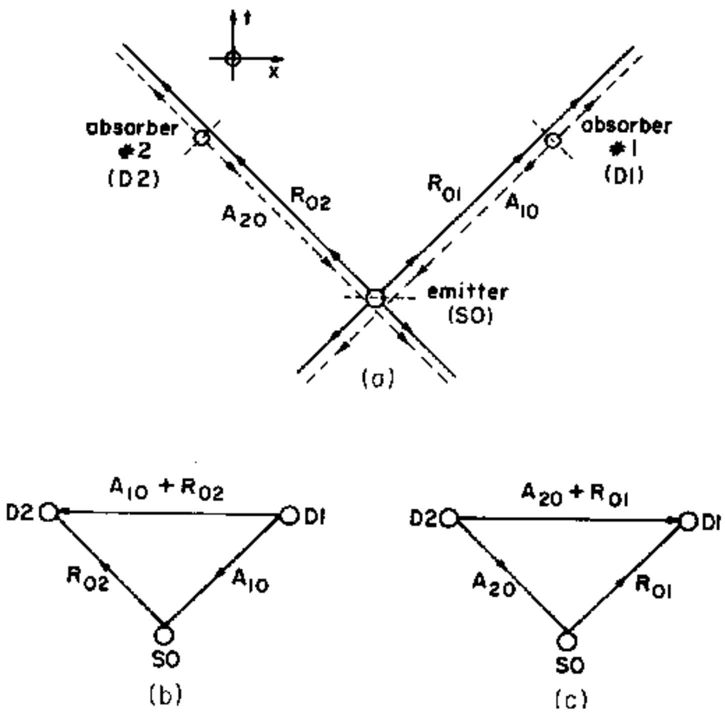


Figure 6 : John Cramer's picture of allowed communication between two observers (absorbers or detectors) in terms of 'Minkowski' space-time diagrams for the case of the 'Freedman-Clauser' experiment. The 45° lines are light-like (at the speed of light on the light-cone) for 'advanced' and retarded waves (arrows down and arrows upwards). Nonlocal enforcement of polarization correlations can occur between D1 and D2 as 4-vector sums connecting them (b,c). The joint transaction occurs along all the lines together and is 'atemporal'. [Cramer, 6]

In spite of the apparent faster than light transfers of quantum information, it is believed that exploitation of nonlocality for controllable signaling is impossible for classical observers ("no signaling proofs").

Following these entanglement (Bell/EPR) tests, Aspect performed a test for the original Wheeler view of a delayed choice experiment. The essence of this test is shown below (e.g., figure 7).

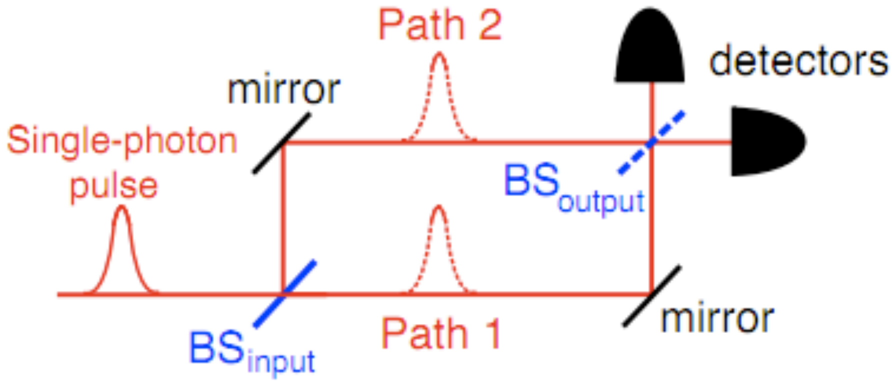


Figure 7: "Wheeler's delayed-choice Gedanken Experiment with a single-photon pulse in a **Mach-Zehnder** interferometer. The output beam splitter BS-output of the interferometer can be introduced or removed (closed or open configuration) at will." The 2006 Aspect realization [5] used a **48 meter** pathway with movable BS output. "The choice between measuring either the open or closed configuration is made by a quantum random number generator, and is space-like separated — in the relativistic sense — from the entering of the photon into the interferometer. Measurements in the closed configuration show interference with a visibility of 94%, while measurements in the open configuration allow us to determine the followed path with an error probability lower than 1% [6]." The traditional initial and final beam splitters have opposite orientations for different phase shifts of reflected and transmitted light so that one detector will usually see just destructive interference (no nominal output).

Feynman referred to the usual two-slit quantum superposition interference experiment for photons or electrons as containing the central mystery of quantum mechanics. However, having two slits close together restricts the types of experiment modifications that can be performed. The Mach-Zehnder interferometer with a large rectangular pathway enables a lot of room between the two interfering paths so that many experiment can be done. Each path can be separately affected by electric fields, magnetic fields, and gravitational fields to vary the phase of that path. This device is increasingly used for modern QM tests. One result is to show that for single particles, particle traits such as charge, magnetic moment, mass, spin, appear to be carried along both paths at the same time.
Later Bell tests include: Geneva 1998 test of entanglement over several kilometers distance. Tests guaranteeing pure randomness for delayed choices. Three particle entanglements. Trapped entangled atomic-ions (Boulder 2001). 2008 18 kilometer detection. Superconducting qubits. Many Zeilinger tests in Austria. As of 2011, physicists have now been able to entangle 8 photons together in a very complicated apparatus on a light table! (the latest record [11]).
Perhaps the most interesting interferometer is a specially carved single silicon crystal that breaks an incoming neutron beam (from a nuclear reactor) into an upper and lower beam and then recombines them again to get interference. This one can see that gravity slows down the phases of the lower beam. A special device can also be inserted into the lower beam so that it responds to the presence of a separate magnetic field that precesses the neutron's magnetic moment and causes a shift in the output interference. The particle property of magnetic moment appears to exist in each path even for just one neutron at a time going through the system. There is interference between the possible upper and lower path-- it really boggles the mind.

References:

1. J.S. Bell: On the Einstein-Podolsky-Rosen paradox. Physics 1, 195 (1964) [one of the most profound papers in science].

2. J.F. Clauser, M.A. Horne, A. Shimony and R.A. Holt: Proposed experiment to test local hidden- variable theories. Phys. Rev. Lett. 23, 880 (1969)
3. S.J. Freedman, J.F. Clauser: Experimental test of local hidden-variable theories. Phys. Rev. Lett. 28, 938 (1972)
4. A. Aspect, J. Dalibard, and G. Roger: Experimental test of Bell's inequalities using time-varying analyzers. Phys. Rev. Lett. 49, 1804 (1982). Free access: http://www.phys.ens.fr/~dalibard/publications/Bell_test_1980.pdf

5. Aspect, et al., "Experimental realization of Wheeler's delayed-choice Gedanken Experiment," free access: 28 Oct 2006 http://arxiv.org/pdf/quant-ph/0610241v1
6. {Cramer_Trans} John G. Cramer, "The Transactional Interpretation of Quantum Mechanics," Reviews of Modern Physics, Vol. 58, No. 3 July 1986, pp. 647-687.
7. {Cramer_Absorb} John G. Cramer, "Generalized absorber theory and the Einstein-Podolsky-Rosen paradox," Physical Review D, Vol. 22, No. 2, 15 July 1980, pp. 362-376.
8. {WebC} John Cramer collected papers on the transactional interpretation of quantum mechanics web sources: http://faculty.washington.edu/jcramer/PowerPoint/Sydney

9. Wolf Prize to Aspect, Clauser, and Zeilinger, February 4, 2010, http://physicsworld.com/cws/article/news/41633
10. Einstein, A; B Podolsky, N; Rosen (1935-05-15). "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" Physical Review 47 (10): 777–780. EPR.
11. Physicists break record for extreme quantum state, Physics World, Jun 22, 2011, http://physicsworld.com/cws/article/news/46331
12. Nick Herbert, Quantum Reality—beyond the new physics, Anchor Books, 1987.

13. Amir D. Aczel, Entanglement [The unlikely story of how scientists, mathematicians, and philosophers proved Einstein's spookiest theory], Plume Book, 2001.
14. "A clever proof, using simple Venn diagrams," B. d'Espagnat, Scientific American 241, 158 (November 1979).
15. Feynman, Leighton, Sands, Lectures on Physics, Vol III, Chapter 11, 1965.

APPENDIX:

Question: Plane polarized photons? Figure 4 shows the cascade from a spin zero state to spin one and then spin zero again. A photon is a spin one particle which carries away the change in angular momentum of the electron states. A spin one photon is either right circularly polarized state |R> or left circularly polarized |L> so that the electric field corkscrews through space in the direction of motion. But the experiments use plane polarized detectors. Any direction of polarization can be written as a combination or "superposition" in the |x> plane direction or the |y> direction [15], and a y-direction plane polarized photon is: $|y\rangle = (|R\rangle + |L\rangle)/\sqrt{2}$. Similarly, a right polarized photon can be written as $|R\rangle = (|x\rangle + i\,|y\rangle)/\sqrt{2}$. If $|R\rangle \rightarrow |y\rangle$, where does the angular momentum go—it has to be conserved. [Ans: probably into the polaroid crystal].

How to visualize the "simplest Bell example" $N(A^+B^-) + N(B^+C^-) \geq N(A^+C^-)$.
Draw a 2x2 square array (2 rows x 2 columns) where the rows are A+ and A- (top to bottom) and the columns are B+ and B- (meaning 'not B' – like in [14]). At the center, draw a circle or center a diamond so that the inside is C- and the outside is C+. Label all little areas from upper right as a "counter-clock-wise" b, c, d for the outer C+ and then repeat around another circle for the inner C- d,e,f,g. Then the above inequality is area "e + a + f + g" >" f + e" – obviously true. If counts are proportional to the appropriate partition of areas, then counts also follow. But it assumes independence (a tall boy doesn't bias towards having blue eyes).
The Scientific American Picture uses: $N(A^+C^-) + N(A^-C^+) + N(B^+C^-) + N(B^-C^+) \geq N(A^+B^-) + N(A^-B^+)$.
Or: "f + e + c + d + f + g + a + d" ≥ "a + e + c + g" , or after canceling, just $2(f + d) \geq 0$. Again, obviously true.

BENEATH QUANTUM MECHANICS

DAVID L PETERSON

Abstract. Apart from a proliferation of new possibilities, only minor progress has been made over the past eight decades in the arena of the interpretation of non-relativistic quantum mechanics. We are impatient for resolution but acknowledge that is unlikely in any near future through proper channels. Therefore, a new proposal is presumptuously ventured here projected from presently known clues listed below. The "Sub-Space-Web" Interpretation is built upon omnipresent quantum fields and includes the following: It is proposed that the wave-function ψ has a rich content that can carry information for momentum, energy, mass, entanglements, spin and polarization, but also coordinates with the quantum fields to carry "particle properties" such as charge, magnetic moment, quantum numbers, and knowledge for complex assemblies such as atoms and molecules. This is assisted by a Vacuum of space-time that is also more intricate, active and capable than usually believed and which essentially forms a highly complex quantum communication network. The physics at the level of quantum superpositions includes an elaborate series of trial simulations of possibilities. Processing by the Vacuum might have a random component but is also complex enough so that each quantum collapse could be an essentially emergent phenomenon. Relevant information exchange between an emitter and absorber require multiple "back-and-forth" communication in time for the production of a completed transaction which is then an element of the classical world.

1. Introduction:

The disciplines of mathematics use terms subject to formal logic and can do so because they are very carefully defined. In contrast, Nature is the owner of real definitions for physics. Our statements of physical terms are iterated and improved by our evolving but imperfect human understanding. Physics texts generally fail to provide definitions of key terms (particle, wave, charge, field, time). And it appears that no text on quantum mechanics provides a suitable definition of its own discipline.

What is quantum mechanics? Assembled and summarized dictionary type definitions may provide a beginning of understanding by saying something like:
Quantum mechanics is a mathematical machine that can predict the behaviors of microscopic particles based on the century-old proposals that their energy and angular momentum may be restricted to discrete amounts called quanta. In quantum systems, matter and energy have aspects that can be both particles and waves. Quantum ideas aided explanations of blackbody radiation, the photoelectric effect, Bohr atomic orbitals, the existence of

Date: 15 March, 2011, paper updated to 3/23/14.
email: davepeterson137@gmail.com. Boulder, Colorado.

discrete packets of energy and matter, the uncertainty principle, and the exclusion principle. 'Quantum mechanics can be regarded as the fundamental theory of atomic phenomena.'

And, with slightly more depth, quantum physics is a collection of models of physical phenomena which use the mathematics of "Hilbert space" to make operational predictions for the outcomes of laboratory experiments. It is a probability theory using the concept of probability amplitudes and the 'Born rule' that probability is the square of the magnitude of the complex amplitude. These concepts may be extended not just to microscopic systems but also to some macroscopic systems. Physical phenomena in quantum mechanics has an action which is of the order of the Planck constant, h. It uses a set of postulates relatively unchanged from the 1920's and 1930's enabling calculations which agree with statistical measurements of quantum systems. It is a mathematical system lacking a consensus on the interpretation and reality of its mathematics and terms.

This paper presents a presently unconventional perspective on the physical reality that may operate in the world of quantum mechanics. Some of the views given here are not necessarily new but are only rarely entertained seriously. Over the past few decades, there has been a revival of interest in the interpretation of quantum mechanics [1], and many new and clarifying experiments have recently been performed [2]. But, there is still no current consensus on interpretation, nor does it seem that there is likely to be one for many decades. It may even be said that there is a growing degree of lack-of-consensus. The purpose of this paper is to propose a platform while acknowledging that it is premature. Most of us probably won't live long enough to see the ultimate theory, but deduction from currently known facts can guide us towards a solution. One aspect of the "Sub-Space-Web" may be that the quantum world between particle emission and absorption is initially a "simulation" of the superposition of all possibilities 'as if they were all *real*.' This sort of processing is continually and naturally performed by the 'machinery' of the Vacuum of Nature. This is not just mathematics, and this level of physical reality is not that of the classical world. It is quite different from the ordinary world and needs its own concepts and language. To state this difference, we first discuss the possible nature of the cosmic medium and problems with some current interpretations of the mathematics of quantum mechanics. We then refresh our knowledge of some suggested tools to assist the new view: the 'Transactional Interpretation' of John Cramer [14] and the alternative 'Feynman Path Integral formulation' of quantum mechanics.

A major problem in trying to interpret quantum mechanics is the traditional bias that concepts and convention must refer to measurements and classical concepts. But, restricting words only to classical and measurement terms when thinking about the underlying mechanisms of quantum mechanics may be like looking for lost keys only under streetlights. It is likely that the quantum world is strongly different from the classical world and ultimately needs a language in its own terms rather than our terms. As an example, suppose for a moment that the Transactional Interpretation has some validity. Then what we call the wave function, ψ, might more correctly be an "offer wave" from an emitter to an absorber. For the case of photons, a photon is its own anti-particle so that an acknowledgement wave, ψ^* could travel backwards in time from the absorber to the emitter thus forming the beginning of a transaction with strength $\psi^*\psi$. Then the major perenial question, "Is the wavefunction real?" is similar to the question, "What is the sound of one hand clapping?" There is a "reality," but in this example it is clearly not a "classical reality;" and there is a big problem with using the word "real." This problem is also seen for the increasingly popular "Many Worlds" – what sense does it make to ask how real is a separate Everettian universe in the multiverse? Here we wish to talk about the nature of the Vacuum substrate and the sub-quantum world of superpositions of possibilities and wish that there were special yet acceptable words for these concepts away from terms used for classical reality with classical biases. Many of the words we use could be placed in quotation marks or italics or clarified with a phrase like "real in the sub-quantum world" (perhaps 'QuReal'). Similarly, a claim such as "a photon has no properties until the time it is measured" presumes that time only moves forward – another classical bias. Perhaps quantum time has some different properties from classical time in the sense of being forced only forwards.

The traditional definition of 'real' includes being a physical entity having properties differing from the 'ideal' and existing independently of mind and human observation. A macabre example might be that Galileo's finger in the Florence Museum exists there as a physical entity before we actually look at it. Someday, perhaps its DNA will be examined to show more details about its history differentiated from other fingers and their owners. The word 'ideal,' on the other hand, suggests existence in the imagination or as an archetype. In a sense, the primary concerns of physics merge these two words together into a more appropriate and perhaps higher and more valuable 'reality.' Each physical particle is an example of an identical invariant form without differentiating properties from ideal. All of the universe's electrons are the same, each of it protons is the same, each of its ground state gold atoms is the same. It is convenient but perhaps naive to think of an elementary particle or excitation as a special type of vibration of tiny unique and identical 'springs' of a universal medium. The existence of identical particles has important consequences. If two particles have trajectories that cross, then perhaps not even Nature knows which particle came from where.

Note that the physical laws and constants of Nature are universal and invariant over space and time. Mathematicians generally believe that their basic theorems and concepts are also pre-existing Platonic Forms which are discovered rather than invented, and many of these concepts apply to Nature. To some large degree, mathematics can be abstracted from physical reality because the nature of physical reality is expressible in the language of mathematics. Formerly, physical entities were said to 'exist' because they had mass. Now we might say they exist because they have energy equivalence (e.g., $E = mc^2$ and $E = h\nu$). Photons and electric fields exist because they have energy even though their mass is zero. A present concern is whether <u>information</u> also has some real existence. Quantum information exists in the quantum world in a different fashion than classical information exists

in the classical world. Quantum information may not be expressible directly in terms of energy.
This paper elaborates further on the problem that language and unstated classical assumptions may have blocked progress in interpreting quantum mechanics.

2. Problems with Quantum Mechanics

Since its conceptual and mathematical formulation in the 1920's, non-relativistic quantum mechanics (NRQM = QM) has had amazing empirical success and has never encountered an experimental counter-example. QM in turn led to relativistic quantum electrodynamics (QED, 1948) which is regarded as the most successful physical theory of all time. Quantum Field Theory (QFT) in general has had its ups and downs in success and fashion from the early hope of the 1930's to confrontation with infinities to successful renormalization in QED to confrontation with strong interactions to the success of quantum chromodynamics (QCD, 1973) and electro-weak unification (1967-1983) and now to its current struggle with gravity, string theory and SUSY and GUTs. But it is at least accepted as an effective theory of Nature. It was initially hoped that a theory of everything would help to explain QM, but so far string theory just assumes quantum mechanics in its formulation.

Despite the successes of QM, QED, and the "Standard Model" QFT unification, there is no present consensus on the **interpretation** of the mathematics of quantum mechanics. The initial success of the Copenhagen Interpretation ('CI') which dominated quantum mechanics for decades is somewhat an accident of history dependent on the charisma of Niels Bohr, the positivism of Heisenberg, the intimidation of de Broglie by the mistaken genius of Pauli (1927), and the abstract formulation of von Neumann and his mistaken "proof" that hidden variable interpretations are invalid (1932). The early de Broglie and Bohm (1927, 1952) 'pilot-wave' interpretation has been growing steadily in popularity and is viewed as a satisfactory interpretation of QM. 'CI' is now gradually eroding against many other self-consistent and apparently legitimate competitors.

Now there are conferences every year on the foundations of quantum mechanics that discuss interpretation, problems, fundamental questions, the meaning of Bell's theorem, EPR dilemma, and paradoxes in QM. One observer of these meetings[1] noted [11] , "You will find all of the religions with all their priests pitted in holy war − the Bohmians, the Consistent Historians, the Transactionalists, the Spontaneous Collapseans, the Einselectionists, the Contextual Objectivists, the outright Everettics, and many more beyond that. They all declare to see the light, the ultimate light." These factions largely add new concepts

[1]− a 'Quantum Bayesian' who doesn't believe any of these popular interpretations: 'Quantum mechanics has always been about information' and 'the quantum state is solely an expression of subjective information' and 'represents observers' personal information, expectations, degrees of belief.' 'Probability is a measure of our presupposition.' [12]

onto the old QM: for example, actual trajectories, universal wave function, objective collapse mechanism. What quantum mechanics needs, if possible, is the ability to tell a story understandable to a high-school student. Plain words should not only interpret but add value such as forming a foundation from which the mathematical theory can be derived. The previous template is Einstein's relativity which provided the 'relativity principle' and constancy of the speed of light essentially as basic axioms. And then for general relativity there is the simply stated principle of equivalence (accelerating elevators versus gravity). And of course there is John Wheeler's statement that "Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve." There has to be "a dynamic interplay between storytelling and equation writing. Neither one stands alone...' [11].

Non-relativistic quantum mechanics should dovetail with relativistic QFT; but conflicts with interpretation exist, and development of interpretation and philosophy of quantum field theory is still primitive. The meaning of ψ in QM is somewhat incompatible with its meaning in the relativistic theory, so it difficult and unclear to say whether QM or QFT is presently more fundamental. QM discusses "particles" already in existence. Some claim QFT is a theory of unlocalizable particles and others as just fields [8]. 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. The relativistic wave-function is a functional of quantum fields rather than function of particle space-time coordinates. Psi is not a probability amplitude as in usual QM but rather operators which create and destroy particles in various normal modes. QFT treats fields as the knowledge embedded in the Vacuum of how to make any particle providing that adequate energy is available to do so [6]. Some say that even in usual QM there is really 'no evidence for particles' [7] [8].

Many physicists accept that QM works and works well, and some can also live with a philosophy of "no-interpretation" (or "shut up and calculate") or the old Copenhagen dogma. Some of these are the finest minds in quantum mechanics. For example, experimentalist Anton Zeilinger says, "I don't hesitate to declare my preference for the Copenhagen interpretation" [52]. And there is ongoing awareness that worrying about interpretation has been and still is a generally unproductive exercise with few tangible successes. The intentional avoidance of interpretational discussions in QM textbooks has saved generations of physicists from this unproductive worry (but has also left them with the impression that although strange and abstract, the original QM formulation is without problems). Nevertheless, from an early age, a goal of physicists is to understand Nature, and that has been impeded by the opaqueness and confusions of quantum interpretation.

In addition to being minimalistic or 'austere,' the standard 1927 Bohr-Heisenberg Copenhagen interpretation has also been imperfectly defined and internally inconsistent [9]. Its basic postulates certainly include: a) The wavefunction ψ can be a superposition of allowable possible states, b) The Born statistical interpretation where probability $P = \psi^*\psi$

(which has always seemed mysterious along with the fundamental existence of **probability amplitudes**), c) Heisenberg uncertainty, d) contextuality (coupling of the microscopic with macroscopic, complementarity, wholeness, 'inseparability'), e) the state vector representing "our knowledge" to aid an understanding of "collapse" (or alternatively that ψ gives a complete specification of what can be known), f) avoiding discussion of microscopic 'quantum reality' (Heisenberg positivism), and g) time evolution of ψ proceeds unitarily via the Schrödinger equation until its "collapse" via measurements. Final collapse of the wavefunction is just a separate rule with no explanation being offered. Functionally, the first three postulates may perhaps be most important. Although it is generally claimed that Born's rule and that wave function collapse have to be added as separate axioms in quantum mechanics, there are still some deterministic approaches from which they can be viewed as derivable [56].

A lesson from Bohr is that experimental setups and detections are understood to be classical with the quite different realm of quantum reality operating in-between. But the boundary between micro- and macro- is vague and has never been well defined. If they indeed 'exist,' particle paths can be superposed, and accelerated electrons need not radiate. But the idea of state information or "our ignorance" has always seemed incompatible with the certainty of the interference of alternatives. Also, Copenhagen infers that there is no 'reality' below measured reality and also that ψ is in a state of all possible superpositions until the time of measurement. But, it is an experimental fact that there has never been a measured result where an object is seen in two states simultaneously, and results must be single eigenvalues of single eigenstates. Bohr acknowledged that there is a difference between the quantum and classical worlds but wished to only discuss each in terms of classical concepts. This avoided many interpretational problems in early history. Bohr was the father of the *operational interpretation of QM* and one of the fathers of the *information interpretation of QM* [61]. But his views changed over time so that there are several Bohr interpretations. The Copenhagen interpretation says that "A wave function provides a complete description of an <u>individual</u> quantum system" (rather than statistical properties of an ensemble of similarly prepared systems as in the *ensemble interpretation*) [2].

Of course, there are many other alternate stated selections of postulate sets for QM: quantum 'states' are rays in Hilbert space, observables are Hermitian operators, measurement 'projections,' identical particle statistics (which originates in QFT), spectral decomposition, complementarity and correspondence principles. CI QM has definite macroscopic outcomes but microscopic super-positions of states. It has deterministic unitary evolution but random Born rule outcomes. Bohr himself never talked about the collapse of wave packets because he didn't consider them to be physically real.

CI is mathematical and abstract and intentionally discourages anyone from looking deeper – no microscopic reality exists by itself. CI doesn't believe that it is possible

[2] which is also called the *statistical interpretation* of quantum mechanics.

to talk about deterministic hidden variable theories prior to measurement. But the deBroglie/Bohm theory is a counter-example self-consistent non-local hidden variable **deterministic** model with particles possessing locations and velocity at the same time along their trajectories [10] and with some degree of 'weak measurement' support [5]. The old realistic, causal, and deterministic Bohmian or "pilot-wave" view is not dying out over time but has five-fold growing references since 1970 with citations currently near 140 per year [3]. Bohm doesn't require the "collapse" or projection postulate nor self adjoint operators, and the Born rule is largely derivable rather than being a postulate. Bohmian "collapse" *is* the selection of a trajectory [4]. There is now talk that "weak measurements" can expose these trajectories [58] [5]. As for positivism, philosophers abandoned it several decades ago with physicists lagging by a decade. [6]. Note that the statement, "the wavefunction contains all knowledge that can be known about a system" itself violates positivism. Some believe that the first-quantization QM is physically inadequate and conceptually needs to be replaced by the "second-quantization" of Dirac and Feynman– perhaps with a more appropriate name like "occupation number formulation" since a quantum field is not a quantized wave function.

Experiments over the past few decades have solidified the mysterious nature of QM whose puzzles might have been previously ignorable: non-locality (although there are still some dissenters), no local-realism, clear entanglement [1] [see definition at end], four-particle entanglement with no common overlapping past [2], Mach-Zehnder interferometry tests, interference of single neutrons [26] and even single complex molecule "buckyballs," orbital angular momentum of light (OAM, up to many $\hbar$'s), quantum computing, quantum teleportation [2], various "Aharonov" test effects, and decoherence.

Perhaps the hardest mystery of quantum mechanics is the apparent appearance of what would be called actual particle properties in multiple locations at the same time. This means more than the wave frequency and wavelengths representing energy and momentum in a wave-function but also quantities like charge, particle type, quantum numbers, spin, and magnetic moments of particles (as in a neutron crystal interferometer). One way for this to happen is for each path to have these attributes as just being Nature's coding of

[3]Yet, neither deBroglie/Bohm nor Cramer interpretation could be considered mainstream; both have a small percentage of followers [64].

[4]Despite its renewed popularity, the Bohm interpretation was intended only as a preliminary step to show that 'independent actuality' was possible in quantum mechanics. Bohm wasn't fond of it and felt that it didn't go deep enough. His quantum potential, Q(x,t) [or 'information potential'] represents experimental conditions and depends on the quantum state of the whole system. This agrees with Bohr's requirement to account for the whole experimental arrangement [67].

[5]Although consistent with QM, one objection against Bohm is that in the relativistic domain it needs a hidden preferred reference frame − currently against the spirit of general relativity.

[6]Philosophers don't accept logical positivism anymore, but many physicists still tend to. This affects their concepts of meaningful versus meaningless questions and causes some antagonism between the views of physics and philosophy. Einstein advocated logical positivism until about the time when Heisenberg insisted on positivism for his quantum mechanics.

information attributes representing particle characteristics. Real classically particle characteristics would then only materialize at selected end points or transformation points of paths. In-between, in a deeper quantum reality, is the processing of information for the **sub-quantum world**. To elaborate further, one of the generally accepted Copenhagen postulates is that before an observation is made, a quantum object simultaneously exists in all of its possible states. A measurement then resolves this superposition by collapsing the wave function into a definite but randomly selected state that can be seen by the classical world. Landau said that an atom has no definite or specific attributes until it interacts with something else that has. It has no position and has no speed until measured (David Bohm could disagree). It is a difficult question to ask what an electron is doing in an atom or what a nucleon is doing in a nucleus. If a nucleus is viewed in a Pauling fashion as stacked balls of nucleons; then because of the high density in nuclei, there is very little room between nucleons for any motion to occur. Yet it is known that nucleons undergo complex motions (orbital angular momentum, independent particle properties, Fermi gas modeling, spins and shells). Real electrons as particles may not exist in an atom. It seems more appropriate just to refer to all the wavefunctions instead. And for the nucleus, it seems that each nucleon is in multiple locations at the same time spread throughout the body of the nucleus. A visual mind considering objective particles is subjected to a horrible strain.

A motivation for the present note is providing a presentation of a framework that does not lead to a highly distasteful "Many Worlds Interpretation" (MWI) for QM. Steven Weinberg diplomatically refers to the branching of the world into vast numbers of histories as 'disturbing' [49]. MWI is viewed as a 'trap' with an increasing following to the point of becoming mainstream [7] It is a logical possibility but is unaesthetic and unnecessarily wild beyond plausibility and contributes little to understanding in terms of mechanisms. It is a consequence of thinking about the quantum sub-world using traditional classical terms and concepts and taking the Schrödinger equation and particle attributes too seriously. It is unable to explain the Born rule, $P = \psi^{*}\psi$. Historically 'reasonable' approaches to quantum interpretation have largely failed so that possibly unreasonable ones like "Many Worlds" are being contemplated and even believed. The 'simulation interpretation' is an alternative to that, and classical "particle reality" would materialize as the result of final hand-shaking agreements between sources and selected receivers.

"Many Worlds" says that the possibilities in a superposition actually all become real but do so without collapse via a splitting of the universe for each quantum event. Every MWI quantum possibility actually happens somewhere in some "universe;" but the world we think we live in only sees single selected outcomes. So, the Schrödinger Cat would be alive in one universe and dead in another split off parallel universe. And 'there is a universe where Elvis is still alive' (Guth). Some of the latest incarnations of MWI include actual

[7] Following near 18% popularity in polls between 1997 to 2011 [64].

realizations of all possibilities in the infinite multiverse of eternal inflation. A line in the sand might say that this sort of talk is sad and has gone too far. The conversion of some to MWI may result from the frustration and desperation of confronting what quantum mechanics seems to be without being able to contemplate other and non-traditional possibilities. Most physicists have not been exposed to the myriad proposed 'interpretations.' A recent alternate view is provided by George Ellis [45]. He says that because of the many layers of reality by size and complexity, laws pertaining to a lower level emerge into other relevant laws at each higher level. The behavior of lower level elements depends on the context in which they are imbedded into the higher levels. Examples of levels include: particle physics, nuclear, atomic, chemistry, materials science, geology, astronomy, and cosmology. Non-linearity is involved in each level transition so that the linear unitarity of QM at low levels gets combined in non-linear ways into the higher levels. This is relevant to the infamous problem of the 'Heisenberg cut' between the (micro) world of QM and the (macro) classical world in which measurements occur. The applicability of QM at higher levels is strictly limited. Then concepts such as Schrödinger's cat, the wave function of the universe, and Everett's Many-Worlds interpretation are challenged.

There is growing appreciation of the depth and power of the Vacuum of our real universe. But it hasn't yet grown to the power and awareness it might deserve and might provide the basis of an alternate explanation to quantum mechanics. What appears to be "Many Worlds" could be the Vacuum's 'self-simulation' of a set of possible quantum transactions. All possibilities are encompassed by the ψ "offer-wave," and all portions of the offer-wave are "physically processed" at the sub-quantum level of reality. Each local Vacuum processor asks and calculates, What If the incoming wave were a real physical particle, how would its physics go?

An interpretation of quantum mechanics should: explain what a wavefunction or state vector is, try to explain 'collapse' of the wave function, state a selection principle for final event versus randomness, discuss reality below the level of classical observers, and explain 'contextuality' $-$ the dependence on a final absorber and on the influence of more classical levels of reality on the context and boundary conditions of the more quantum levels [45]. MWI lacks mechanism and cannot deal with entanglement. Another problem with many worlds is that "the theory is not only explicitly observer-dependent, but even worse there is an element of arbitrariness concerning the choice of basis that is used for the decomposition of the universal state vector into worlds [8]. Another minimalist interpretation of quantum mechanics, the 'ensemble interpretation,' only addresses the wave function as representing a statistical ensemble of results with no reality for individual transactions. The 'simulation-interpretation' turns this on its head and suggests that each single event results from actual sub-quantum processing much of the possible statistics together.

[8]'The pointer basis problem,' Craig Hogan, ArXiv 1204.5948 28 Sep 12.

One of the defining characteristics of quantum mechanics is the so-called 'wave-particle duality.' 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.' 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." [8]. Einstein was a supporter of "fields are all there is view of physical reality fields are states or conditions of space. This is the modern view." "Quanta are countable, but they are spatially extended and certainly not particle-like." And Casey Blood argues that, "Wave-particle duality arises because the wave functions alone have both wave-like and particle-like properties [7].

If only waves exist without particles, then much of the problem of superposition becomes natural. All possible paths become one widely extended wave. Path integrals largely become Huygen wavelet analyses in disguise. Momentum is deducible from mean wavelength, energy is mean vibration, mass is just semi-localized rest frequency. Mass in general is 'bound energy (apart from the Higgs field acting on 'elementary particles). Localizing is only possible subject to an uncertainty principle deducible from Fourier transforms of waves. An atom is somehow a set of cohesive waves.

3. Clues Leading Beneath present Quantum Mechanics:

- Sub-Classical: QM has its own special properties such as spin, uncertainty, entanglement and superpositions; the quantum world is different from the classical world. The Born rule even suggests that ψ lives in the 'square root of reality' which includes the hypercomplex number systems. $\Rightarrow$ Cease restraining QM descriptions and formalism to only familiar Classical terms and analogies.
- Constants: There exist a large number of precisely defined universal physical constants and the ability to create physical particles from the Vacuum. $\Rightarrow$ The Vacuum of Space-time must have complex hidden micro-structures enabling an effective memory of discrete physical parameters.
- Math: Physics is highly mathematical and obeys universal physical laws. $\Rightarrow$ In some sense, the Vacuum is a "Cosmic Computer" and has a set of effectively stored values everywhere.
- Information: Neutron interferometry, Buckyball interferometry, and Mach-Zehnder experiments show that "particles" can exist in different places at the same time. $\Rightarrow$ Realize that wavefunctions refer to information about particles rather than particles themselves. "Particles" may not even exist; waves and fields may be everything.
- Waves: QM and QFT use waves and fields where $p = \hbar k, E = \hbar\omega, m_o = \hbar\omega_o/c^2$. $\Rightarrow$ Waves and Fields are somehow real; and wavelength, frequency, and rest frequency are spacetime <u>codes</u> for momentum, energy, and mass.

- "Simulation:" A single photon can travel through a glass prism with the proper index of refraction and usual degree of bending. ⇒ Prior to final output, the quantum world is a trial simulation over all possibilities (e.g., photon E-field effectively interacting with myriad electrons in the glass).
- New disciplines of Quantum Information and Computing: ⇒ The evolving ability of quantum computing technology reflects the underlying capability and complexity of spacetime and its 'Vacuum.' [9]
- Path Integral or Sum Over Histories formulation/interpretation is considered fundamental ⇒ myriad network connection paths are available through space-time.
- Zero-Point Fields: Calculations of the cosmological constant Λ from zero-point energies fail profoundly. ⇒ Base wave-functions only live in a simulated world where energy is not classically real until quantum transactions have been completed.
- Time Reversal: QM exhibits contextuality and dependence of the wavefunction on its measuring apparatus ⇒ Cramer "back-and-forth" communication in time seems to be required for this and also for explaining entanglement. Time can behave differently in 'the square root of reality" sub-quantum world. This is also needed for the functioning of teleportation.
- Wheeler's delayed choice tests: Last instant decisions can be made to observe a wave or particle aspect well after a photon began its journey. This makes more sense if the detectors and choices can communicate their presence backwards in time to the source [10]
- For multiple particles, the Schrödinger wave function is on 'Configuration Space' of 3N-dimensions. This is entanglement of particles. Along with delayed-choice experiments and entanglement swappings, this seems to also imply a 'bi-temporal' quantum communication network between 'particles.'
- Contextuality: A quantum event is a triplet of a state preparation, a processing, and a state measurement − all together. It is the system as a whole that counts rather than any of its parts. This suggests an ongoing communication covering both initial and final states (a sub-space-web covering the cases of both individual particles and entangled particles).

Elaborating further:

A goal is to propose a foundation for a speculative but plausible quantum mechanical world not directly perceivable by classical thinking and measurement. Since most previous approaches are found lacking despite many decades of trying, it may be philosophically

[9] 'A 2011 poll of physicists said that, 'Quantum information is a breath of fresh air for quantum foundations." [64]

[10] The first clean test of Wheeler's delayed-choice experiment using true single photons was presented in 2007 and also used quantum random decisions on the interferometer being closed (giving interference) or open (giving so-called "which-way" photons). Linearly polarized single photons from single nitrogen-vacancy centers traveled 48 meters in the interferometer allowing time for last instant measuring choices [68].

appropriate to state what we think we believe and start again from the beginning:

(1) Inadequacy of conventional Language: It is a tradition in physics to only work in areas that can produce publishable and hopefully verifiable papers (string theory perhaps excepted). When dealing with the 'actual' quantum micro-world, appropriate words have a somewhat new and special meaning that is unfamiliar to us from a classical perspective. What is a Wave? Wavefunction or State-Vector? Particle? Energy? Vibration? Space, time, vacuum, simulation, dimensions, momentum, velocity, speed, distance, ... QM lives in a different world from that of classical mechanics. We have to be very careful in using common words because they may block or derail rather than enable understanding. We could almost attach a 'Q-' for 'quantum' or perhaps a 'psi' ($\psi-$) before each familiar word (the sub-quantum world of **superpositions of possibilities** is "psiland," and ψ is psi-real). Needed words don't yet seem to exist, but also language depends on the interpretation model being used. Since the 'correct' interpretation is unknown (and possibly even unknowable), any special language is provisionary. The definition and meaning of the 'wavefunction' or 'state' is the biggest problem in quantum interpretation. Zeilinger says that a wave-function has no properties until the time of actual measurement [2], and he assumes time progresses only forwards. But the Cramer transactional interpretation would say that psi already knows about the type and results of measurement at the time of emission from anti-photons moving backwards in time so that communication is atemporal. With this understanding, sub-quantal psi waves could be considered as non-classical real or 'QuReal'; it is just a different level of reality. Conventional 'psi' might be better defined as an 'offer-wave.'
As another example, in classical physics, mass times velocity is momentum. But Bohm would disagree with that understanding and say that the quantum potential makes up the disconnect.

(2) Mathematics: Nature is highly mathematical and obeys a finite set of laws with a finite set of constants. Why? What does it mean? And we also now know that a Vacuum is 'not nothing.' It might be assumed that this is connected with physics being a mathematically defined process. Much of useful mathematics somehow pre-exists in Nature. And much of this mathematics involves hypercomplex numbers that are classically unfamiliar to us. Make serious use of the fact that both quantum mechanics and classical mechanics obey mathematical laws and do so to great precision. The 'unreasonable effectiveness of mathematics' in physical science has always been a mystery and would seem to imply a mathematical construction to the universe and its Vacuum. Inside the sub-quantum world, there is a transition between physical information and physical reality. The information includes all the physical constants, conserved values, and precise laws of Nature duplicated endlessly. A single quantum event might have intricate sub-detail. At present, the

only visual picture we have of the underlying mathematical complexity of Nature is something resembling the existence and duplication of hyper-dimensional knots attached to each point of some three-dimensional space mesh − something like the Calabi-Yau space curled up at each space-time point of our classical space [55].

(3) Psi-Waves: Assume that Waves have some sort of 'actual' existence. The waves of the wave-function are carriers of momentum and energy, p and E, information, but the waves live in a quantum world which may not bear much resemblance to classical expectations. They are "real" but in quotation marks (or 'QuReal'). It could be that Feynmans path integral is an accurate representation of what occurs − not just a formulation but also an interpretation. That would mean that instead of considering broad wave-fronts, it may be more correct to consider all of the paths or threads through space time that lead to the common phase wave-fronts. The complexity of the Vacuum could enable it to behave as a functioning network or part of a cosmic computer.

(4) Assume that Energy is indeed (or at least encoded by) some sort of vibration (rest mass-energy, moving mass-energy, 'pure' energy − obeys $E = h\nu$). Vibration of what? Note that the units of Plancks constant 'h' are joules-second or energy per hertz of vibration (or units of momentum per wavenumber or unit of angular momentum). As in relativity, energy and time transform similarly; and momentum and wavelength go together. These are measures of wave *concentration* in time and space. In Nature's units, Planck's constant can be set to one.

(5) Complexity: Micro-Nature is incredibly complex and can go down deeply into sizes. This complexity could emerge the phenomenon of collapse and could enable information to be carried as internal quantum numbers like charge. It is possible that extra dimensions are also involved. It is certainly difficult to explain the many different elementary particles capable of being plucked from the Vacuum. Complexity could also enable the duplicated storage of physical laws and constants.

(6) Entanglement: Schrödingers second 1926 paper on quantum mechanics showed that his wavefunction psi for several particles is on a multidimensional configuration space. He originally hoped that his waves represented some real oscillation but was finally forced to state in his June, 1926, fourth paper of wave mechanics that $|\Psi|^2$ is a kind of weight-function in the configuration space of the system. The wave-mechanical configuration of the system is a superposition of all configurations allowed in the mechanics of points. New students tend to avoid this conundrum by focusing on single particles or realizing that the 2-particle hydrogen atom wave function in a configuration of six coordinates can be restated using relative coordinates into (discarded) center of mass coordinates and the useful three electron coordinates. But the configuration space for multiple particles was an early realization of "Entanglement" (a term finally coined by Schrödinger in 1935 regarding

the EPR paper) [see Entanglement discussion at end]. Now we have experiments for 'entanglement swappings.' In a recent example of 4-photon trajectories, two photons become entangled such that the remaining photons do not share coexistence at the same time. One photon is created and measured prior to the other being created and measured still giving full quantum correlations [60]. Without backwards and zig-zag communication in time, this would be truly weird. Configuration space, entanglements, delayed choice experiments (contextuality) are clues pointing towards bi-temporality quantum communications.

(7) Other Possible Clues:
zero point states do not gravitate.
$\psi^*\psi$ is actually taken and used in chemical bonding without the presence of observers (covalent bonding depends on a non-classical Born enhancement of electron density between nuclei).
The Big Bang creates space from nothing.
Space-time is deformable.
Tunneling really occurs.
Quantum output appears to be truly random.
Superpositions cease upon measurement.
Identical particles exist and affect statistics.

4. The Vacuum:

Our view of the Vacuum of space-time has evolved over the past century and can play a role in the interpretation of quantum mechanics. The Vacuum is now seen as much more complex, detailed, active and intricate than previously believed. It is 'not-nothing,' it is a "thing." It is not just a state symbolized by $|0\rangle$; it is more like a "machine" that facilitates physical processes. It is increasingly seen as something much more than "empty space" and is assumed here to now be endowed with some as yet undetermined structure and interconnectedness enabling memory and processing capability for physical constants, physical particles, and physical laws. I've always called this 'new vacuum' the embodiment of the ancient Greek concept, "Plato's Form Heaven." And things like electrons and electromagnetic fields are perfect examples of Plato's spaceless and timeless invariant forms (much more so than he could have ever imagined). Each type of quantum field is a 'Form' or archetype of Nature. Of course, talk about the nature of a universal medium is still intangible; and there has been prolonged bias over the past century from the demise of the "luminiferous aether" in 1905 (but Poincare and Lorentz continued to value the aether concept). Properties of the successors to the luminiferous aether include: General Relativity with a dynamic space-time metric responding to changing energy and radiation sources and being able to carry gravitational waves and having space being dragged around rotating black holes (i.e., the 'Einstein ether' $\equiv g_{\mu\nu}$), a variety of quantum fields occupying space-time in which Nature is essentially made of fields, condensates in the Vacuum, and background fluctuations and zero point energies. The 'fields' are the 'forms.' The use of the word "vacuum" is another example of the inadequacy of existing language in describing quantum reality. About the only current term for a "substance of space-time" is still "aether." But this word is highly unfavorable, has old baggage, and is even taboo in formal physics. Einstein's and Dirac's use of the concept of a relativistic ether 'lacking immobility' never caught on. Rather than placing "vacuum" in italics or quotation marks, the use of capital Vacuum is used here until a better term is found. Its high importance also justifies capitalizing the word.

Frank Wilczek also decided that a new word was needed which is broader and more relevant to physics and QFT than the old terms of vacuum, space-time, aether, plenum, substance, or world-stuff [46]. He uses the word **'Grid'** as a multilayered, multicolored cosmic superconductor encompassing quantum fluctuations, a unified superconducting condensate, a weak superconducting Higgs condensate, Einsteins metric field, the dark energy cosmological constant grid density, and chiral symmetry-breaking condensate consisting of quark-antiquark pairs. This recognizes that general relativity is very much an ethereal theory of gravitation. The boiling background of virtual quantum fluctuations is detected via screening for QED and antiscreening for QCD. Grid superconductivity gives masses to particles created by weak bosons. Particles are localized disturbances in the Grid. Some add that the smoothly distributed cosmic black-body background (CMB) is also a modern version of aether with a locally preferred frame corresponding to the expanding cosmic fluid. Wilczek's picture is now encouraged by the experimental finding of 125 GeV resonance appearing to be the standard model Higgs boson (CERN-LHC, 4-July, 2012).

Brian Whitworth [50] has a 'Virtual Reality Conjecture' that the usual physical reality is a digital output of quantum reality processing which in turn is the transforming of information. He uses many analogies from computer science such as screens and pixels and rebootings. His 'screen' producing output is similar to Wilczek's 'Grid' underlying physical reality − a network of dynamically connected nodes. There is no empty space, and conservation laws just conserve processing of certain attributes. A photon is viewed as a program distributed over the grid as packet instances, a spreading processing wave encompassing all Feynman paths. Although photon instances travel all possible paths, they tend to arrive in straight lines while not actually traveling in them. He perhaps goes too far in suggesting that time is processing cycles, collapse is due to a processing overload, and fields are network properties.

The Vacuum is a complex medium with properties. It was known long ago, for example, that free space has a characteristic impedance of 376 ohms. This can also be expressed as $Z_o = |E|/|H| = \mu_o c$ where the speed of light $c = 1/\sqrt{\epsilon_o \mu_o}$. Fundamental physical parameters include c, Planck's constant $\hbar$, electron mass m_e, electron charge q_e, the Newtonian gravitational constant G_N, and perhaps the permitivity of free space ϵ_o . These values should be built into our Vacuum. Forty years ago, one would have said that basic constants included things like the mass of the proton, m_p, but that is now semi-derivable from lattice-QCD. Other basic quantities like the Bohr radius are derivable: $a_o = 4\pi\epsilon_o\hbar^2/me^2$.

The basic parameters of the current "Standard Model" of particle physics (SM) include: the masses of the e, μ, τ leptons, the u, d, s, c, b, t quarks, the CKM mixing angles (θ_{12}, θ_{23}, θ_{13} and CP violating phase δ $-$ for which the Nobel Prize was awarded in 2008), θ_{QCD} vacuum angle, and the Higgs values of μ and λ (for a Higgs potential $V = \mu^2\phi^2/2 - \lambda\phi^4/4$). Alternate choices may include the coupling constants such as the strong constant ($\alpha_s$$-$ but it really isn't constant and changes with energy). Also note that the electromagnetic fine-structure constant changes with energy too: $\alpha_e = e^2/(\hbar c 4\pi\epsilon_o$ in S.I.) has the value 1/137 which increases to 1/128 at the Z_o mass scale. Additionally, or alternatively, we could include the the mass of the neutral Z^o, the weak boson Higgs vacuum expectation value $v = 246$ GeV, the neutral Higgs mass near 125 GeV ($m_h = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}$), and the Fermi coupling constant $G_F = v^2/\sqrt{2}$. With recent developments, one should now also add neutrino masses for ν_e, ν_μ, ν_τ (or, ν_1, ν_2, ν_3) along with their mixing matrix as well (seven additional parameters). It could be claimed that these values are stored in the Vacuum to be pulled out anywhere as needed. Also consider dark matter and the cosmological constant Λ_o or dark energy. If there is physics beyond the standard model such as models having super-symmetry (SUSY), then there may be additional fundamental physical parameter values as well. An estimate for the information content of the Standard Model and its parameters is 'about 200 bits, hardly one line of text' [44]. While unimpressive in our world, universal storage in the Vacuum might seem to require a complexity requiring extra dimensions. These may possibly be the extra dimensions of string theory where a string is a one-dimensional object moving through a given background space-time.

So, the Vacuum is no longer seen as a simple thing but is filled with propensities, structure, information and fluctuations. The frontiers of speculative theory suggest loops of space, tiny strings, spin-networks, and the mathematical-like processing of gauge theories. The idea of spin-networks goes back to Roger Penrose in 1971 with re-discovery by Rovelli and Smolin later. Loop Quantum Gravity (LGC) deals with very fine fabric or network weaved of tiny loops. Spin models include networks of strings and "string-net condensations" which claim that photons and electrons can be considered as emergent phenomena [3]. Space-time and gravity itself could emerge from underlying geometric relationships in LGC [51]. The universe might be formed from quantum computation.

As a recent example of the content of space-time, consider particle-antiparticle colliders producing what might be called "pure energy" which then in turn can output myriad possible output particles of precisely defined types apparently emerging out of the Vacuum itself. Since the earth rotates and orbits, the real 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. Distant light and cosmic rays in astronomy also indicate the invariance of physical laws and constants throughout the universe. There is a beautiful recent plot released by CERN LHC showing quark-mesons produced by the vacuum as seen by increasing total mass of di-muons, $\mu^+\mu^-$ (see Fig. 1 at end [40]). The spectra of events per GeV begins at left showing spikes 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 boson near 92 GeV. These particles are spewed forth when the Vacuum is stimulated.

The Vacuum is highly complex, and its processes are highly complex. As an example of this, consider that long computations in lattice-quantum-chromodynamics (L-QCD) so far can yield the mass of the proton to about a few percent accuracy; so QCD "explains" the proton and the neutron and the general excited states of hadrons and mesons. But, despite the acceptance of QCD as a quantum field theory, it has yet to be able to derive any basic nuclear physics. Even the existence of the smallest compound nucleus, the deuteron, is currently computationally beyond the abilities of lattice-QCD − tritons, helions, and alpha particles should be even harder. Calculations for these and beyond are believed to require "exa-flop years" of computer effort (10^{18} floating-point operations per second sustained for years on the best super-computers in the world) [11]. That sort of computing power lies well in the future − beyond the present tera-flop-years and peta-flop-years. But the Vacuum of Nature does all this with ease everywhere all the time. There are probably many more fundamental particles or excitations of the Vacuum than previously perceived. It must have layers of reality more than several orders of magnitude smaller than current thresholds of testing (e.g., lengths smaller than 100 GeV and living in 'zepto-space'). It must have high information content at tiny scales. Quantum computing is suggesting ultimately more computing power in the micro-world than for large-scale electronic computers. It is not uncommon to refer to the universe as an "information processing system," a quantum computer, and a cellular atomaton [4]. But these are poor and puny terms for what Nature actually does.

5. Cramer's Transactional Interpretation (TI):

John Cramer published his major review on "The Transactional Interpretation of Quantum Mechanics" in 1986 [14]. This interpretation is based on the 1945 Wheeler and Feynman electromagnetic absorber theory of radiation that used not just standard retarded waves but also their consistent coupling with advanced waves moving backwards in time from all possible absorbers [15]. Cramer's application to quantum mechanics provided an intuitive understanding of nonlocality and of the EPR paradox where advanced waves enable a verifier for quantum mechanical transactions. A quantum event is a "handshake" between an emitter and an absorber. How a particular handshake is selected is not given and might be random and unknowable. Unlike any other interpretation, the previously mysterious concept of entanglement and multi-entanglements of particles become intuitively clear in TI. In particular, TI can explain the results of the Freedman-Clauser experiment

[11] However, computational algorithms are also improving over time

where polarized correlated light violated the Bell inequality. Although generally ignored in the 1980's, Cramer's ideas are very appealing and some preference for his scenario has been slowly growing [17] [21]. One author recently stated, "TI deserves serious and open-minded reconsideration" [22]. Advanced and retarded solutions are also discussed for any massive particle as well.

Cramer is far from alone in his focus on a time-symmetric quantum mechanics. Other approaches towards this view include studies by Griffiths, Unruh, Gell-Mann, Hartle, Heaney [63] and Schulman. One approach is to "use the insight from Cramers Transactional Interpretation where the time-reversed Schrödinger equation (SE) was assumed to be on equal footing with the ordinary SE [38]." But Cramer did not use symmetric BCs and still invoked collapse. The previous work most mathematically similar to Cramer is Aharonov and Vaidmans 'Two State Vector formalism' [34]. And the early Feynman-Stückelberg interpretation of antimatter is that it is just corresponding matter particles traveling backwards in time. Relativistic wave equations such as that of Dirac or Klein-Gordon allow both negative and positive energy solutions; so for consistency to the non-relativistic case, there is no reason that the Schrödinger type equation should only represent just forward time processes. Historically, the SE was derived by considerations of the relativistic Klein-Gordon Equation (KGE). But, 'one can reduce the KGE to a single SE only by artificially discarding half of the solutions − the so called negative energy solutions [38].' And Schrödinger purposefully dropped one of the solutions in the process. So, there really should be **two** Schrödinger equations given by:

$$-\frac{\hbar^2\nabla^2\psi}{2m} = +i\hbar\frac{\partial\psi}{\partial t} \quad \text{and also} \quad -\frac{\hbar^2\nabla^2\psi}{2m} = -i\hbar\frac{\partial\psi}{\partial t} \tag{1}$$

Photons are their own anti-particles (as are the neutral pion boson π^o, Z^o and also the Higgs, H) described by the operator of conjugation, so their backwards wave function is ψ^*. Anti-particles derive from QFT and their back in time possibility is generally ignored in standard QM. It is also possible that neutrino's are their own anti-particle, and testing is underway to explore this possibility [39].

Rather than just postulating the Born rule for $\psi^*\psi$, Cramer "explains" it naturally as the cascade of forward offer wave ψ evaluated at the absorber locus together with the backwards in time verify wave ψ^* evaluated back at the emitter. All possible absorbers have a ψ^* wave back to the emitter, but only a selected absorber at a time has the relevant active Born rule probability. Although $P = \psi^*\psi$ is intuitively clear in TI, it is also possible that $P = |\psi|^2$ just happens to be a rule imposed by the coding in the cosmic computer. Collapse is viewed as a completed transaction but with language that is not yet well developed, 'the emitter responds to the echo and the cycle repeats until the response of the emitter and absorber is sufficient to satisfy all the quantum boundary conditions' [14]. [12] Future elaboration of this atemporal "pseudo-time" development is desired. That opens up the possibility that the first pass, the 'offer wave' is a wave but that the final transaction pass is the 'particle.' The classical world would deduce both happening together in time (a 'wave-particle duality'). But they would be consecutive in pseudo time. Cramer is compatible with the "de-Broglie/Bohm" interpretation/formulation if a Bohmian "particle" is identified with a "transaction". So, in a sense, Cramer's TI 'explains' Bohm's particle with a guide wave. One apparent difficulty with Bohm is that his particle isn't much of a particle– at any intermediate time it only possesses location and velocity and not much else. But in these models, particleness only occurs at or after the completed transaction. There is no need for observers in TI − just absorbers to give a 'transition to definiteness' [45]. Note that a Vacuum capable of arranging a transaction doesn't really require the physical existence of a particle traveling in-between a source and absorber anymore than an electronic money transfer needs to have coins and dollar bills actually travel from a giver to a receiver. An agreement is made that the source of the money just subtracts a given amount from his reserve, and the receiver adds that amount to his. The money is conserved like energy and momentum and other needed values are conserved.

An understanding of entanglement is especially attractive in Cramer's view. Intuitively, it would seem that entanglement requires the appearance of nearly instantaneous communication between two distant entangled objects. Penrose also stated, This flitting back and forth in time is precisely the kind of thing that 'quanglement' [his name for quantum entanglement] is allowed to do" [27]. One example is the usual emission of two photons by one excited atom producing entangled polarizations. The case of four-photons with transferred entanglement is also easily visualized using Cramer's back and forth 'W-shaped' zig-zag pathways through space-time. But there are spin examples too. Mathematically, let an example be:

$$\psi = \frac{1}{\sqrt{2}} \left(|0\rangle_A \otimes |1\rangle_B \ - |1\rangle_A \otimes |0\rangle_B \right) \tag{2}$$

So, if A = 'Alice' measures a '0,' then the state of the system collapses to just the first term. Distant system B = 'Bob' is altered or "unveiled" by A performing a local measurement.

Further progress in TI seems to require an improved understanding of **time**. Time has always been a fundamental mystery, and it is easy to state a long list of unanswered questions. **What is Time?** [13] When is Now? How is time flow coordinated across space, and is time a result of coordination between the separate elements of space? Does space itself contain a fundamental clock? Why the Arrow of time? What is time reversal? Does time have an origin? Is inertia and time flow due to a Mach's principle mechanism?

[12] Some believe that this 'hierarchy of transactions' is not needed and can be replaced with a block-universe interpretation [57]. TI can explain the paradoxes of 'delayed choice,' 'contingent absorbers,' 'interaction-free measurement,' and 'the quantum liar experiment.'

[13] *["Time is Nature's way of keeping everything from happening at once." :)]*

What is time dilation? Is time an absolute external parameter for particle dynamics? How can time evolve from quantum theory? Is quantum time intrinsically different from classical time? Is atomic time or universal time the ultimate clock? What are better words to express repeated 'back-and-forth' communications in time– the 'hidden' time (network time)? It is possible that classically observed forward progressing time is the result of sets of myriad back-and-forth exchanges and resolutions. The future consists of possibilities involving quantum mechanics, while the past is decided and classical. In a sense, then, time is the advancing step function between future QM possibilities and previous classical resolutions. Prior to the Cramer Interpretation was a 1964 time-symmetric quantum formalism by Yakir Aharonov called the "Two-State Vector Formalism" [TSVF] [36]. This uses a prepared initial state $|i\rangle$ and a 'post-selected' final state $\langle f|$ where the final state propagates backward in time. In a sense, this scheme begins where Cramer leaves off. After the transactional negotiating and selection emergence, then there is a final confirmation wave going backwards in time. I see these two ideas as dovetailing with other (although there is no suggestion of this in the literature). The TSVF idea seemed to stay in historical limbo until a later paper on "Weak Measurements" was developed from it [35] [33]. This is now a very hot topic with much discussion and experimentation. [see Definition at end].

The transactional interpretation of quantum mechanics has failed to become a dominant interpretation partly due to 'back-and-forth' time transactions being a little too wierd for conventional physicists. Part of the problem is a powerful bias from the classical world "Arrow of Time" impacting belief in backwards in time sub-quantum communication. The recent views of George Ellis [65] may help here. He proposes that the arrow of time originates from the initial singularity and cascades down from cosmological to micro scales through a hierarchy of structures – a 'top-down' influence. At the sub-quantum micro-scale, equations are symmetric in time. The origin of time asymmetry is due to the initial and/or final conditions, e.g., the prepared source and the macro-sized detector impose the arrow of time. But sub-quantum communications are free to go forward or backwards in time.
John Cramer's contributions to TI diminished while he practiced more traditional physics (and ostensibly retired). Many of the newer contributions now come from Ruth Kastner [23]. She offers a "possibilist" TI stressing "ontologically real possibilities existing in a pre-spacetime realm" [24]. The 'PTI' formulation includes offer-wave possibilities as physically real potentiality or 'potentia.'

The name 'Reality' is not restricted to just familiar classical reality but also to these interfering sub-quantal actions as well. No one is offering a name for this non-classical reality, and one is really needed for the prepared/emitted quantum state prior to its detection/absorption. Actualized events are rooted in unactualized possibilities.

"The only valid criterion for choosing one interpretation rather than another is how effective it is an aid to our way of thinking about these mysteries and on that score Cramers interpretation wins hands down." "Cramer's interpretation is very much a myth for our times; it is easy to work with and to use in constructing mental images of what is going on, and with luck at all it will supersede the Copenhagen Interpretation as the standard way of thinking about quantum physics for the next generation of scientists." A major virtue of TI is that it 'tells a story' to go along with the mathematics.

Cramers retro-causation was also recently advocated in the American Journal of Physics [4]. "It is concluded that Einsteins spooky actions may occur in the past rather than at a distance, resolving the tension between quantum mechanics and relativity and opening unexplored possibilities for future reformulations of quantum mechanics."

Note that even in the Bohm case, it would seem that some back-and-forth communication is necessary merely to establish the relevant wave-function prior to using it to create a quantum potential and establish a relevant final pathway.

PROBLEM: Although it may seem that a Cramer transaction is the most intuitive explanation for the Born rule, there are some glaring exceptions. One is that the atomic orbitals for covalent atoms form $\psi^\star\psi$ and use it for effectively enhanced electron spatial density between atoms for bonding without the presence of absorbers. Another is that electric and magnetic fields have energy density proportional to E^2 and B^2. Having energy density is a traditional requirement for declaring existence. It now seems likely that the vector potential field $A(x, y, z, t)$ is primary but that Nature actively takes derivatives of this field to form E and B fields. In some sense, Maxwells theory of EM is the relativistic quantum theory of a single photon in disguise (with deviations from Maxwell theory coming from multi-particle effects). So having the Vacuum form energy as the square of EM fields is largely equivalent to forming $|\psi|^2$. Both of these cases suggest that Nature processes the Born rule by itself throughout the vacuum. It is likely that gravity always interacted with 'matter in a quantum way. We indeed have gravity pulling on neutron wavefunctions over separate pathways in the famous "COW" crystal interferometer experiments. In reverse, these wavefunctions may gravitate backwards. Space-time seems to know the Born rule.

Finally, there is the very important issue of "photons that never end" $-$ photons that never terminate on an absorber. And this does appear to happen. Most of the universe's history of optical, infrared, and cosmic microwave photons (COB, CIB, CMB relic backgrounds) are still traveling freely through space after many billions of years. And with an accelerating expansion and thinning of the cosmos, they won't interact in the far future or even ever. In absorber theory, this presents relevant incomplete 4π absorption. We know that this radiation phase does contribute to the energy content of the universe and also to net gravitation. And because their density loses energy as scale factor a^4, they lose energy as they travel in transit. We can't talk about the individual existence of these photons, but they certainly have a collective existence in the primary way that existence matters: having energy and gravitating. So far, there is little support for the idea that there might be some characteristic length after which photons must localize. What seems to remain is the need for each region of space to be aware of all the probability amplitudes that are in it and make an estimate of how to gravitate and then effectively really do it. This means that quantum mechanical wave functions in any region of space have effective energy density depending on the frequency of each photon and their amplitude squared:

$$\rho_{\psi's} \propto \Sigma_{all\ photons} |\mathcal{A}_i|^2 \nu_i \tag{3}$$

This suggestion goes beyond what most people would concede for the processing power of the Vacuum. But it seems to be a logical new requirement.

6. Feynman Path Integral, an Alternate Formulation of Quantum Mechanics

The Feynman Path Integral ("PI") or "Sum over Histories" formulation of quantum mechanics is considered by some to be more fundamental than the Schr̈odinger wave formulation or the Heisenberg matrix mechanics formulation [14]. One can derive the Schrödinger equation from the path integral [13]. PI is more concerned with 'transition probabilities' than with the more unobservable states or wavefunctions. It is also not just restricted to non-relativistic QM but aided in creating quantum electrodynamics (QED) and guided other quantum field theories as well. Quantum field theories are encoded in the Lagrangian, and this mechanism yielding phase is also key to PI.

Path Integrals have extended use for naturally occurring fluctuating line-like structures like particle orbits, general relativity, polymers in solutions, vortex lines in superfluids, defect lines in crystals and liquid crystals and even for the study of financial markets [32]. For the standard example of traversing space-time from one point to another, one considers all possible pathways (histories) and sums up their possibilities using their phases along the paths. These include non-classical paths such as going backwards in time, being kinky or curlicued, taking excursions into distant space and back again. The basic postulates for the path integral formulation are:

(1) The standard "Born" postulate that the probability for an event is given by the squared length of a complex number called the "probability transition amplitude".

(2) The probability amplitude itself is given by adding together the contributions of all the histories in configuration space.

(3) The contribution of a history to the amplitude is proportional to $\exp(iS/\hbar)$ where $\hbar$ is reduced Planck's constant, and can be set equal to 1 by choice of units, while S is the action of that history, given by the time integral of the Lagrangian along

[14] Formulations of Quantum Mechanics include the standard or 'canonical' formulations called the matrix formulation, the wavefunction formulation, the pilot wave formulation, path integral, and density matrix formulation. One could also include the 'phase space formulation,' Hamilton-Jacobi, second quantization, and the variational formulation. Other views suggest an 'algebraic formalism,' (e.g., C*-algebra of observables) and a 'quantum logic formulation' (or quantum propositional calculus)

the corresponding path.

In order to find the overall probability amplitude for a given process, then, one adds up, or integrates, the amplitude over the space of all possible histories of the system in between the initial and final states, The path integral assigns all of these histories amplitudes of equal magnitude but with varying phase. The contributions that are wildly different from the classical history are suppressed by the interference of similar, canceling histories The quantum action is generally just appropriately discretized classical action. Feynman showed that this formulation of quantum mechanics is equivalent to the canonical approach to quantum mechanics, when the Hamiltonian is quadratic in the momentum. An amplitude computed according to Feynman's principles will also obey the Schrdinger equation for the Hamiltonian corresponding to the given action. Dyson showed that PI for QED was equivalent to the field theories of Schwinger and Tomonaga. Note that Feynman was unusual in advocating the primacy of particles [8]. Feynman paths and Feynman diagrams can be recast into a 'fields-only' picture. It should be allowable to replace an ensemble of paths with an ensemble of cascading Huygen wavelets instead. They should be equivalent pictures.

When the electromagnetic field is included, the Feynman path integral for an electron incorporates an additional change in phase due to the vector potential of the field.

$$\Delta\varphi = \frac{e}{\hbar}\int A\cdot dx \simeq eA\frac{\Delta x}{\hbar} \tag{4}$$

This is also a key equation in Gauge Field Theory. In the special case where we only care about the gauge function $\chi(x,t)$ of the gauge freedom $A' \to A + \nabla\chi$, the integral is trivial and only depends on end points rather than being path dependent. This equation (4) is also core to the important Aharonov-Bohm effect $-$ but with non-gradient fields (such as the A field outside a solenoid). Without them stating so, it must be understood that for this to happen, the full value of electric charge (or the information for the electric charge, e) must somehow "effectively" be carried along each path over many paths at the same time. A similar interpretational difficulty occurs when a neutron passes through a crystal Mach-Zehnder interferometer and has its phase change due to an EM interaction in one or both of its paths. The full particle magnetic moment (or its knowledge) must be transported along and processed along each path. In a sense, the 'particle properties' of charge and magnetic moment exist in separate paths at the same time. It is evens worse when a single buckyball passes through two slits at once and then interferes with itself. It makes one contemplate the existence of "Many Worlds" with particle properties propagating fully through each history. The alternative to this actualization may be that only knowledge or information or 'wave' is actually transferred and processed with a final 'particle property' materialization occurring only at collapse.

A 'Network View' of Feynman paths would consider pathways of little line segments as 'quantum-real.' Such a view is consistent with Huygen wavelent rebroadcasting of waves from each point in space-time so that the wave-view is consistent with the network-view. For long paths, only the monotonically advancing leading Feynman paths would end up contributing to the leading wave amplitudes so that all of these views are compatible.

Although path integrals may be fundamental, unfortunately they are also difficult to work with. Only a few trivial examples can really be handled analytically. This realization is a barrier to the introduction of path integral methods into basic quantum mechanics courses. Here is one example of the difficulty: By 1972, Feynman was embarrassed by not being able to solve the hydrogen atom in a Coulomb field using his path integrals and challenged a physicist named Hagen Kleinert to attempt it. In 1978, Duru and Kleinert found a key idea that time has to be replaced by a new path-dependent pseudo-time along with new square-root coordinates. But it wasn't until 1989 that Kleinert realized that new concepts like gradient torsion and curvature contributions were also needed. A Duru-Kleinert non-holonomic transformation is required to take the path integral of the H-atom into that of the harmonic oscillator, but this transformation introduced curvature and torsion [32] . It was unexpected that the path integral of the Coulomb system could become a laboratory for testing non-euclidean space torsion. Just because these calculations are difficult for us doesn't mean that they are difficult for Nature to perform. QCD, for example, is exceedingly difficult for us to do; but, as mentioned before, Nature does it every tiny fraction of a pico-second and continues this processing for the lifetime of the universe and over 10^{80} particles or more.

Let a 'thread' be a sub-quantum path from a single source to any point in space-time (essentially a real Feynman path for the path integral formulation or sum over histories). Different paths are superpositions of possibilities. There are many thousands of threads per event and they only interact in terms of amplitude and phase with phase mainly varying by the mass and momentum of a particle. But also let each thread carry the attribute of relevant charge, spin, identity, and other quantum numbers. Although replicated many times, these values do not add physically or mathematically − they are information attributes. But each thread attribute can participate in interactions with external fields to alter the phase present over each path. Most threads will interfere destructively, and the leading ends will be coherent with constructive interference. Interference results in only having contributions from the stationary points of the action give histories with appreciable probabilities. So, although 'threads' may be highly kinked, they can conveniently be pictured more as straight rays. So, when we say that single particles exist in separate places at the same time (e.g., in interferometry), we only mean that particle attributes appear in each thread set and are able to alter local phase. Actually it is worse than this. Particle behavior exists in many places at once. The sub-quantum world involves the processing of information in space-time and by thread prior to detection of quanta in the classical world. The effective materialization of electron charge in chemical bonding is curious. Pictures of RHIC collisions of extremely energetic lead nuclei show myriad rays advancing radially over a spherical shape. These rays re like the coherent classical appearing paths of threads of possibility.

7. "The Square Root of Reality":

Take the Born rule seriously as having sub-quantum-real ('Qureal') wavefunctions needing to be 'squared' to become classical candidate entities. Classically recognizable probability is given by $P = \psi^*\psi$ where psi lives in a new subworld resembling the pulling apart of classical reality into two "square-root" parts. The 'star-root' (or 'square root') of reality suggests a subworld represented by complex or hypercomplex mathematical representations. So, electron spin as classically real fails to agree with observation, but quaternions or gamma matrices seem to fit needs better. All discussions of the Born rule go from wavefunction to probability. Here we wish conceptually to go backwards, from classical to sub-quantum and glance at the mathematics relevant to that world.

As a first example, one occasionally used representation of a photon ('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 $\psi = \sqrt{\epsilon_o/2}(E \pm icB)$ [15]. The 'star root' operation is of course not unique and not well-defined, it is intended to only be heuristic: Star-root ${}^*\sqrt{Prob} = P^{*/2} = \psi$. Another example is considering the Dirac equation as the star root of the Klein-Gordon equation, $\partial^2 + m^2 = (i\partial - m)(-i\partial - m)$, so $Dirac = (KG)^{*/2} = i\partial - m$.

$$i.e.,\ \hbar^2\partial^2 + m^2c^2 = (i\hbar\partial - mc)(-i\hbar\partial - mc) \ \rightarrow\ i\hbar\gamma^\mu\partial_\mu\psi - mc\psi = 0 \tag{5}$$

In this new square-root world, 'spin' can exist and do what quantum spin does. The Dirac equation emerges electron spin and electron magnetic moment. [16]

In the classical world, time can only flow forward. But, in this new world, quantum information can travel backward in time. This is recognized by the relatively new field of quantum computing for transmission of qubits [17] , classical bits, and 'entangled ebits' which seem to have backward in time pathways to work [54] (but not mentioning John Cramer). The equations of the quantum world don't imply a direction of time; the direction for time is emergent in classical reality. The Cramer interpretation can work because it deals with entities that exist in the sub-quantum world.

In how many ways does it make sense to take the 'square-root' of a real-numbered classical world? The mathematics of this world obviously includes the complex numbers, $\mathcal{C} = \{z : z = x + iy\}$ where complex $i^2 = -1$. Complex numbers are essential in quantum theory. But we can and do also go much further into the hyper-complex

[15] e.g., see ArXiv reports by Bialynicka Birula

[16] The Dirac equation is $i\hbar\partial\psi/\partial t = \hat{H}\psi = [c\alpha \cdot p + \beta mc^2]\psi$ where we let $p \rightarrow p + eA$ and $p = -i\hbar\nabla$. Or, $i\gamma^\mu\partial_\mu\psi - m\psi = 0$ with partial derivative going to 'covariant' derivative, $\partial_\mu \rightarrow D_\mu$.

[17] Named somewhat as a joke after the ancient 'cubit.' There are also 'ebits' for two entangled qubits 7.

world. This includes the quaternions , $\mathcal{H} = \{q_i = i\sigma_i\}$ [18] with basis $\{1, i, j, k\}$ such that $i^2 = j^2 = k^2 = ijk = -1$ with three complex dimensions (Hamilton, 1843). And the octonions, $\mathcal{O}$ [53], have seven complex bases (1844) $-$ and all the possible combinations of these algebras. John Baez said that as a family analogy, "The quaternions, being non-cummutative, are the eccentric cousin who is shunned at important family gatherings. But the octonions are the crazy old uncle nobody lets out of the attic: they are nonassociative." The 2×2 Pauli spin matrices are related to the hypercomplex quaternions $\sigma_i = \pm i q_i$ where ${\sigma_i}^2 = +1$. The Dirac 4×4 γ- matrices are a subset of $\sigma^i \otimes q_i$, e.g., $\gamma^o \equiv \beta = I \otimes \sigma_3$, $\gamma^i = \sigma^i \otimes i\sigma_2$, ${\gamma_i}^2 = -1$. The world of Dirac matrices can introduce the concept of left or right 'handedness' or 'chiral fields' ψ_L or $\psi_R = \frac{1}{2}(1 \pm \gamma^5)\psi$, where $\gamma^5 \equiv i\gamma^o\gamma^1\gamma^2\gamma^3$. Other cases of interest are covered by topics such as: Lie algebras [e.g., $su(2)$ = span of vector quaternions], Clifford algebras [19] [e.g., $\gamma' s = Cl_{1,3}(\mathcal{C}) = \mathcal{H}(2)$; space-time $= Cl_{1,3}(\Re) = G_4$], biquaternions $(\mathcal{C} \otimes \mathcal{H})$, bioctonians $[\mathcal{C} \otimes \mathcal{O} \sim E_6]$. Electrons spin operators live in biquaternion space (non-relativistic quantum mechanics on Pauli matrices). And relativistically, we can go deeper down into Dirac space.

Instead of just complex conjugation (or *'starring'*) , higher dimension spaces can use matrices with conjugate transpose (or adjoint or 'Hermitian conjugate' or 'Hermitian Adjoint,' or just 'dagger' for physicists $A^\dagger = (\overline{A})^T = \overline{A^T}$, or for mathematicians just A^*). So, the analog of $\psi^*\psi$ in NR-QM may be $\psi^\dagger\psi$ or $\overline{\psi}\psi$ where $\overline{\psi} = \psi^\dagger\gamma^o$ is an 'adjoint spinor.' Having the wave-function and sub-quantum-reality live in the hypercomplex world makes it clear that Bohr's restriction of classical terminology to this world will be very challenging. But I only have a dim memory that anyone at all has ever referred to the 'square root of reality.' The world of quantum mechanics is very different from our classical world, so we must be very careful and cautious about its terminology. In particular, the word 'real' is so mis-used that it should almost be banned.

8. "Simulation Interpretation of Quantum Mechanics":

Thinking about the meaning of quantum mechanics as the simple and literal interpretation of the mathematics is nearly a last resort amounting to "thinking the unthinkable thought:" that the microworld is an internal 'simulation' of possibilities. Simulation is usually implemented by computer modeling; but in this case, it would be 'real in the quantum world.' This is probably also an old thought with the habit of being commonly dismissed as beyond belief. But in the present conundrum, it may be worth reconsidering. The view

[18] My preference is $q_i = -i\sigma_i$ (dp, '76). But there are many alternatives, e.g., S. Wolfram uses $k = \left(\begin{smallmatrix} o & i \\ i & o \end{smallmatrix}\right) = i\sigma_x$.

[19] A Clifford Algebra $C(n)$ is the associative algebra over the reals generated by n anticommuting square roots of -1 $(e_1, ...e_n)$, *with* ${e_i}^2 = -1$, *and* $e_ie_j = -e_je_i$. Intuitively, these are the generalizations of the complex numbers and quaternions [Clifford (1845-1879)]. $\mathcal{H} \sim C(2)$ with two square roots e_1, e_2, $e_1e_2 = -e_2e_1 = q_3$.

here has some similarities to Wilczek's 'Grid' [46] or Whitworth's 'Virtual Reality' [50].

Reasons for believing that the quantum world might be a simulation include:

(1) The previously stated fact that physics is highly mathematical leading one to speculate whether the universe and micro-reality might be a cosmic computer obeying strict laws. At least we now all accept that its substrate Vacuum is 'not-nothing.'

(2) Computers can process alternate and even conflicting possibilities simultaneously [e.g., "What if unemployment rises?" versus, "What if unemployment falls?"]. Superposition of even opposing possibililties is then not so mysterious.

(3) A leading puzzle of QM is how a single particle traversing different paths can have full particle properties (full mass, momentum, charge, magnetic moment, spin or polarization). Obviously, a charged electron traversing a multitude of adjoining Feynman paths cannot have its 'real' charge duplicated for each path because the collective sum of these charges would be infinite. If these are just informational, then actual masses and charges do not have to be duplicated or passed off into many worlds. Mass might simply be a deduction from wave dispersion (for example, if $\omega/k = c = \partial\omega/\partial k$, then mass must be zero).

(4) Calculations of the cosmological constant Λ from zero-point energies fail profoundly. This could be explained by having base wave-functions only live in a simulated world where energy is not classically real until quantum transactions have been completed.

(5) Many branches of splitting worlds could now just be replaced by many concurrent simulations of which only one is selected. For example, David Deutsch [48] suggests that a single photon in flight is really an infinite number of photons expressing all possibilities simultaneously of which only one chosen in the world we believe we call ours. They exist in parallel universes. Isn't it simpler to say that particleness only classically exists after an interaction and the myriad paths prior to that are simulations over possibilities in the quantum reality.

How does a single photon propagate through a pane of glass? Or, how does a photon go through an optical fiber or a glass prism, stimulate its atoms to give the right index of refraction, and then bend its outgoing ray properly?

A large classical electromagnetic wave would interact with myriads of charged particles of the system causing each of them to respond and contribute their own retarded fields to the incoming wave [43]. The wave obeys Maxwell's equations with resulting refraction, reflection, multi-reflection interference, effective slower propagation through the glass due to an index of refraction, and minor absorption and scattering. The resulting output intensity would be decomposed a superposition of these parts. The case for a single 'one-photon' is very similar except that the output is just a quantum of one choice from these parts.

How does this happen? The tiny electric field of the photon gets spread out along possible pathways and interacts with all the charges of the glass making them oscillate and contribute their own resulting electric fields. This is the interesting point − a single photon does this! That has profound implications about photons and about the reality of the Maxwell fields. It is highly inadequate to refer to a photon as just a 'wave of probability.' A photon that is transmitted will have lost no energy due to all this electric interacting unless it is absorbed or scattered. Alternatively, there are many separate Feynman paths with each path mimicking the photon so that the single photon is effectively strongly delocalized yet operating with full information content everywhere. Except for the quantized output, the photon acts like superpositions of little Maxwell waves. An unusual but possible way of thinking is that the photon wave causes a 'simulation' of possible outcomes by the system. Again, this term has inappropriate connotations. One thinks of a simulation by an separate computer − but in this case the simulation is done in place internally by Nature itself, and the simulation is 'real in the quantum world.' The electrons in the glass aren't shaken in a classical sense but rather shaken in quantum simulation for all the 'what-if' possibilities of polarizations and pathways. One often thinks of glass objects as classical bodies. But, in this case of minimal absorption, they are quantum participants which avoid "detection" and decoherence of photon wavefunctions. Perhaps Maxwell equations are really meant to function at this simulation level before they seem to be expressed classically. In the linear approximation, the index of refraction of glass is independent of the electric amplitude − so one cannot state that each path has full amplitude. But, for the Aharonov-Bohm effect, one can make that claim − the full electron charge exists along all paths. Another example is that in modern QM experiments, single photons often traverse tens of kilometers of optical fibre pathways. In this case, different index of refraction in the cladding causes total internal reflection restricting photons to the silica core.

The concept of an effective index of refraction experienced by single particles is not restricted just to light. Neutrinos passing through matter also act as if they experienced tiny values of $n - 1 \sim 10^{-6} - 10^{-20}$ depending on whether the matter is very dense neutron star or light density earth. This is due to weak currents between neutrinos and ordinary matter. The name for this effect is 'MSW' (for Wolfenstein, Mikheyev, Smirnov − 1978 − 1985) [47]. In particular, neutrinos passing through the sun induce enhanced neutrino oscillation. Again, interaction with (possibly light years distance) of "classical bodies" acts quantum mechanically.

Somehow, Nature seems to be able to perform the equivalent of Fourier Transforms (going from waveform in space or time to wavenumber or frequency in space or time). It is not yet clear how this happens, but if it does, it explains the Heisenberg Uncertainty Principle. That is, let a presumed 'particle' have an associated Gaussian probability envelope such that its probabililty 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 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:

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

The case of Gaussian envelopes is optimal and gives equality. Any other waveform envelope profile will give $\Delta x\Delta p > \hbar/2$.

An old question is whether a photon has or even is an electric field. It often acts 'as-if' it does but perhaps the resemblance is merely by simulation. It is more likely that a photon EM field is more like a quantum 'vector potential' $\vec{A}(x,t)$ wave first and that the Vacuum processes this informational wave by taking local derivatives or differences in space and time together to reveal an effective $E = -\partial A/\partial t$ and $B = \nabla \times A$ and then decides how to simulate their effects on a system. For example, if $\vec{A}$ propagates in the x direction with linear polarization in the z direction, then E_z might be $A_o\omega\cos(kx-\omega t)$ and $B_y = kA_o\cos(kx-\omega t)$ the usual crossed wave fields − but they are derived rather than primitive. For the Aharonov-Bohm effect, the A-field is used directly. The A-field might have the meaning something like the 'dragging of electro-magnetic space-time' due to motion of charged sources. In this sense, it would be similar to 'gravito-magnetic' off-diagonal h_{oj} terms from the $g_{\mu\nu}$ metric of general relativity − the dragging of inertial frames.

9. Discussion:

So, how can quantum mechanics be explained?

The **"Sub-Space-Web" interpretation** includes the following ideas:

1. The Vacuum of Space-time is some sort of 'substance' which holds memory of all physical constants and laws.
2. It acts as a quantum computer; and prior to classical output, quantum computation performs a simulation of all possibilities.
3. Sub-space is an atemporal information communication network machine. The path integral uses 'web' paths.

(4) Its entities exist in a non-classical 'square-root of reality' which allows information to be communicated back-and-forth in time without violating the 'no-signalling theorem' of super-luminal communication in the classical world. Quantum communications can appear to be instantaneous and may involve multiple back-and-forth communications. Contextuality is explained essentially by having advanced knowledge of the detector(s) and measuring apparatus. Consciousness and 'Our Knowledge' is irrelevant.
(5) It probably uses added dimensions (for needed complexity) and acts as if it was based on complex/hypercomplex number systems.
(6) The wave-function or state vector is a carrier of coded information which can be decoded at will by spacetime.
(7) The Born rule may be explained by using Cramer's Transactional Interpretation. But it is also possible that it occurs always and everywhere in space-time even without absorbers.
(8) Collapse is not directly addressed (and may involve a truly random computer decision based on the output of simulation).

We lack good conventional analogies for the memory of spacetime. One of the few available familiar (but inappropriately complex) contrived analogies is that of "DNA" or shorter RNA. For life, every cell of a body contains duplicated information for all cells of the whole body. Similarly, each point of space-time Vacuum itself contains the knowledge of all possible particle physical constructs. It is more likely that the information storage knowledge is in the form of complex higher dimension geometrical objects which can be activated in special modes such as a mode for each type of particle. The standard analogy for a quantum field is a 3-D "mattress" or oscillations of springs that can vibrate in modes depending on spring-constants and configurations [42]. A "field" such as the quantum "electron field" may just be associated with these special activations of the pre-existing modes of the Vacuum. These are not just imposed on the Vacuum but are already <u>contained</u> in the Vacuum. The ability of quantum rules to be conserved and processed implies something resembling an "intelligence" to space-time. But one has to be careful about applying terms from common experience to the quantum world, there are very few that quite apply. The Vacuum is a very complex and respectable "machine" obeying strict rules and not some kind of "mind" or "consciousness." Calling it a "cosmic computer" at all scales of size might not be unreasonable. It is hard for us to imagine this complex structure from a Euclidean 3-D perspective. But allowing extra dimensions, complex dimensions, new geometric objects at every point, or knots and loops enable greater imagination and flexibility. Perhaps such added complexity could enable attaching particle properties evoked by the ψ-function without needing actual particles while the wavefunction is evolving. Although the "Vacuum as Computer" is a fairly obvious deduction, it is rarely mentioned. And a typical human would tend not to believe it because it is clear to our direct sense perceptions that empty space is indeed empty. It takes "extended senses" and thinking to begin to believe otherwise (see cosmic microwave universal radiation, neutrino background, zero point radiation, Casimir effect, LHC and tevatron collisions, lattice-QCD modeling, inflation modeling, Bell test experiments, ... etcetera). And of course ourselves and our earth are fairly transparent to neutrinos and WIMPS, almost as if we weren't here at all. In a sense, empty space being empty and bulk-matter not being empty is an illusion.

Communication back-and-forth in Time: Basic QM concepts such as contextuality, entanglement [41], and 'wave functions in configuration space' all seem to imply an interconnected holistic and "effectively instantaneous" communication between associated quantum objects and their interaction with detectors. "Back-and-forth communication" in space and also in time provides knowledge between the parts of a transaction and of their future types of measurements. At least some aspects of Cramer's transactional interpretation must be true [14] or Aharonov "Two-State Vector Formulation" [34]. This view makes entanglement easily intuitive but strains our highly primitive understanding of what time is in quantum mechanics. We are conditioned to think only of forward progressing classical time– but QM requires us to think "out-of-the-box." It could be that there is not just one back-and-forth offer-and-acknowledgement communication but perhaps a "sequence" of a great many of these in each single transaction. Of course, this subquantum time flexibility still does not allow effective classical superluminal information transformation ("No Signalling Theorem").

For Mach-Zehnder Interferometer test pathways using single particles, particle properties such as charge can appear to be expressed fully in each path. This bizarre but rarely stated realization may be a motivation for the Many Worlds belief. One has to decide whether transport of actual physical properties occurs between an emitter and an absorber, or is it adequate simply to transmit information of particle properties and have them actually appear only at end points and obey conservation laws there? Actual charge cannot be duplicated in multiple paths at the same time. It is possible that the wavefunction traversing those paths must itself incorporate or evoke particle property knowledge $\psi = \psi(E, p, m, S, S_z, q, \mu, ...)$. In other words, the old Copenhagen concept of wavefunction expressing information or knowledge may be true. Ideas for the "reading" by spacetime of particle information in the wavefunction are discussed below. Our understanding of the word "particle" is highly biased by a classical perspective and must be something quite different in QM. A "particle" detection result and space-time final transaction location could result in many ways. A 'true' random number generator could arbitrarily select the details of a solution; a final selected path to a detector could involve the actual transmission of a particle or just its knowledge; a "particle" could materialize just at the endpoints of a transaction. It could be that particleness doesn't exist at all– just the apparent transmission of particle properties [7]. The conservation of quantum numbers from origin to detection or absorption is a rule super-imposed by spacetime on transactions. If the wave-function does carry properties normally attributed to particles, then thinking of actual particleness transport from emitter to absorber is almost redundant. No one really understands charge yet; it is an "internal" quantum number with unclear meaning whose derivation lies beyond the standard model (SM).

Assuming that wave-functions represent possible ‘physical realities’ (meaning ‘classical’ rather than a very separately defined ‘quantum reality’) leads to absurdities such as the zero-point estimate of the cosmological constant being excessive by more than a hundred orders of magnitude. It makes more sense to say that quantum properties have no classical reality until full transactions with quantum discrete results have materialized. The quantum linear harmonic zero-point ground state must only have a mathematical existence unrealized in the ’real’ world. Zero point energy is not classical energy and does not gravitate.

The measurements and theories of particle physics have been“Inward Bound” from microscopic sizes to atoms to nuclei to single hadrons to quarks and now to speculations about sub-physics for another twenty or more orders of magnitude in smaller sizes down to the Planck length. The importance of quantum information theory stresses that quantum states represent information that future observers can possess. The common parameters of energy and momentum of a particle are encoded in the wavefunction via frequency and wavelength or wavenumber and perhaps are ‘read’ using local space-time differentiation: $\partial/\partial t$ for energy, $\partial/\partial x$ or ∇ for momentum, and $\partial/\partial\phi$ for angular momentum. Or that eigenvalue approach could be a human construct. Wave density in spacetime is a information code for momentum and energy without assuming immediate existence of particles. “Instead of thinking of particles as colliding, we should think of the information content of the particles being involved in a computation [51].” Like the wave, perhaps neighborhoods of space constantly communicate atemporally so that they can view a spatial wave over extended regions at once. The shape of a wave over a set of many Δx’s can be determined and hence known at each point of space. Electron and neutron magnetic moments are more difficult but could be encoded and derivable in and by the vacuum $-$ perhaps using extra dimensions. The origin and nature of “charge” in physics is not yet established and may involve higher symmetry groups beyond SU(5). But the Vacuum “knows” the Dirac equation– electrons and positrons and quark-antiquark pairs are easily created out of the vacuum and automatically possess their proper spins, charges, and magnetic moments. And it does all this very quickly.

Alternative approaches to interpretation may include the following: It might be possible that particle quantities such as charge and mass appear as quantized or discrete only during measurements due to boundary constraints [37]. Possibly the quantum nature prior to measurement is continuous rather than discrete. As an example, charged particles in electromagnetic fields only experience deflection according to their e/m value so that both e and m could diffuse or spread out together to small values over many multiple pathways as long as e/m remained constant. Indeed, this has always been the appearance of “electron cloud” orbitals in an atom, and some continue to claim that electric charge density per unit volume should be proportional to $\psi^*\psi$ (e.g., the chemical covalent bond). This time-independent stationary-wave view is however inconsistent with the time-dependent Aharonov-Bohm effect. The difference could be that chemical orbitals have continuous reinforcements of their wave-functions. Another remote possibility is that actual particles do fully traverse each of multiple paths but only once-at-a-time in "pseudo-time" (the "back-and-forth" time of Cramer's quantum mechanical time with total duration zero in "real time"). This unexamined view could almost be considered a relative of Many Worlds. The mechanism for a final selection of one path is unknown (see section on "Collapse Analogies" in the Appendix at end). "Collapse" and measurement in general are still huge problems in quantum mechanics. It is hard to imagine the selection mechanisms underlying the collapse phenomenon. One idea may be that emitter and absorber possess a hidden "pointer" and that requiring a match-up the orientations of these pointer arrows may reduce the set of possible transactions.

As an example of space-time computation, "How do fields propagate?" A standard numerical technique for solving Laplaces equation, $\nabla^2\phi = 0$, is to divide up the space between two potential boundary surfaces into cells and iteratively update the value of the potential ϕ in each cell over the average of the values in its neighboring cells until numerical results are stable − using some variant of Jacobi or Gauss-Seidel iterations. For example an $n+1$'th iteration from previous n'th values at 3D x,y,z cell i, j, k may look like:

$$\phi*_{i,j,k} = (\phi_{i+1,j,k} + \phi_{i-1,j,k} + \phi_{i,j+1,k} + \phi_{i,j-1,k} + \phi_{i,j,k+1} + \phi_{i,j,k-1})/6. \tag{8}$$

It is highly tempting to say that this is how Nature also does it - **simple average with respect to neighboring space-time 'cells'** − and does so for every different type of field. One might say that each region of space continually calculates and stores and updates and processes these values. Note that cellular atomata (CA) update cell states in terms of states of neighboring cells. And in Loop Quantum Gravity, neighboring states are also stressed: "Quantum Graphity has the representation of pre-geometry as qubits of adjacency" [51]. The idea above of a regular mesh for spacetime is just a visual convenience; and the idea that each cell carries a real-number potential may also be a stretch. Perhaps this is more like a phase value instead − the modulo 2π of larger range numbers. Then neighboring cells would compare the differing values of the phases as key information.

References

[1] Louisa Gilder, The Age of Entanglement, when quantum physics was reborn, Vintage Books, N.Y., 2008

[2] Anton Zeilinger, Dance of the Photons, From Einstein to Quantum Teleportation, Farrar, N.Y., 2010

[3] Michael Levin, Xiao-Gang Wen, "Colloquium: Photons and electrons as emergent phenomena," Rev. Mod. Phys. Vol. 77, No. 3, July 2005, p 871-879.

[4] Philip Gibbs, "A Universe Programmed with Strings of Qubits," http://arxiv.org/pdf/1106.3029 .

[5] Sacha Kocsis, Boris Braverman, Sylvain Ravets, Martin J. Stevens, Richard P. Mirin, L. Krister Shalm, Aephraim M. Steinberg. "Observing the Average Trajectories of Single Photons in a Two-Slit Interferometer." Science, 2011; 332 (6034): 1170-1173 DOI: 10.1126/science.1202218 first experimental support of de Broglie/Bohm formulation of quantum mechanics. June 2, 2011. [see Correction in arXiv:1109.4436].

[6] Mark Whittle, Cosmology: The History and Nature of Our Universe, The Great Courses guidebook, 2008, e.g., pg. 168.
[7] Casey Blood, "No Evidence for Particles," arXiv: 0807.3930, also see 0912.2985.
[8] 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.
[9] "C.I.J.M. Stuart, "Inconsistency of the Copenhagen Interpretation," Foundations of Physics, Vol 21, No. 5, 1991, pp. 591-623.
[10] Mike Towler Institute Workshop, Tuscany, DeBroglie-Bohm quantum mechanics [Pilot Waves, examples, measurement, nonlocality, relativity, calculations, status, metaphysics]. There are eight pdf lectures available at this site, www.tcm.phy.cam.ac.uk/ mdt26/PWT/lectures/pilot_waves.html/bohm7.pdf .
[11] Christopher A. Fuchs, "Quantum Mechanics and Quantum Information," based on 2001 quant-ph/0106166. Also see: "Interview with a Quantum Beyesian," arXiv:1207.2141v1 [quant-ph] 9 Jul 2012.
[12] N.D. Mermin, "Commentary Quantum mechanics: Fixing the shifty split," Physics Today, July 2012, p 8-10 (and also p 46).
[13] R. P. Feynman, A.R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill Book Company, N.Y., 1965.
[14] John G. Cramer, "The Transactional Interpretation of Quantum Mechanics," Reviews of Modern Physics, Vol. 58, No. 3 July 1986, pp. 647-687.
[15] John G. Cramer, "Generalized absorber theory and the Einstein-Podolsky-Rosen paradox," Physical Review D, Vol. 22, No. 2, 15 July 1980, pp. 362-376.
[16] John Cramer collected papers on the transactional interpretation of quantum mechanics web sources: http://faculty.washington.edu/jcramer/PowerPoint/Sydney_20050723_a.ppt, http://faculty.washington.edu/jcramer/theory.html, http://faculty.washington.edu/jcramer/TI/tiqm_1986.pdf
[17] John Gribbin, Schrödinger's Kittens and the Search for Reality, Solving the Quantum Mysteries, Little-Brown, 1995.
[18] Pavel V. Kurakin, George G. Malinetskii, Howard Brown, Conversational (dialogue) model of quantum transitions, arXiv:quant-ph/0504088v2 14 April 2005 (accessed February 12, 2008); Pavel V. Kurakin and George G. Malinetskii, A Simple Hypothesis on the Origin and Physical Nature of Quantum Superposition of States, arXiv:physics/0505120v1 [physics.gen-ph] 17 May 2005.
[19] Martin A. Ulman, LIGHTNING, Dover, N.Y., 1969 general overview of the lightning process. For high speed video of lightning, see: YouTube.com and Tom A. Warners www.ztresearch.com
[20] SA. Garnier, "Pheromone trail networks in Ants," 9/22/09, http://icouzin.princeton.edu/pheromone-trail-networks-in-ants/
[21] Ruth E. Kastner, "The Transactional Interpretation, Counterfactuals, and Weak Values in Quantum Theory." Studies in History and Philosophy of Science Part B 39 (4):806-818, 2008. Or see, http://philpapers.org/rec/KASTTI
[22] Ruth Kastner, "Why Everettians Should Appreciate the Transactional Interpretation," http://philsci-archive.pitt.edu/5138/1/TIQM_VS_Everett4.pdf .
[23] Ruth Kastner, List of 17 ArXiv articles on the transactional interpretation: http://arxiv.org/find/quant-ph /1/au:+kastner_R/0/1/0/all/0/1
[24] Ruth E. Kastner, The Transactional Interpretation of Quantum Mechanics, The Reality of Possibility, Cambridge University Press, 2012.
[25] R. Michael Jones, Quantum Gravity and Machs Principle, see: http://cires.colorado.edu/mjones/physics/essays.pdf 51.2 MB.
[26] Daniel M. Greenberger, "The neutron interferometer as a device for illustrating the strange behavior of quantum systems," Rev, Mod. Physics 55 #4, October 1983. This article can also be seen at http://hubcap.clemson.edu/ daw/ D_PHYS455/RevModPhys.v55p875y83.pdf.

[27] Roger Penrose, The Road to Reality, A complete guide to the laws of the universe, Alfred Knopf N.Y.,2004.
[28] Wojciech H. Zurek, Decoherence and the Transition from Quantum to Classical Revisited, http://arxiv.org/pdf/quant-ph/0306072v1, (An updated version of PHYSICS TODAY, 44:3644 (1991) article).
[29] Wojciech Zurek, Decoherence, Einselection, and the Quantum Origins of the Classical, arXiv:quant-ph/0105127v3 19Jun2003, pp. 1-52.
[30] Thomas Monz, et. al., "14-qubit entanglement: creation and coherence," arXiv:1009.6126v2 quant-ph 23 Mar 2011
[31] David Kielpinski, "Entanglement and Decoherence in a Trapped-Ion Quantum Register," PhD Thesis, 1996, see: http://www.nist.gov/pml/div688/grp10/upload/dkthesis.pdf
[32] Hagen Kleinert, Path Integrals in Quantum Mechanics, Statistics, Polymer Physics and Financial Markets, World Scientific, 4th Ed. 2006 (now fifth edition) 1547 pages.
[33] Y. Aharonov, S. Popescu, J. Tollaksen, "A time-symmetric formulation of quantum mechanics," Physics Today Vol. 63, no. 11, 27-33, November 2010.
[34] Yakir Aharonov, Lev Vaidman, "The Two-State Vector Formalism: an Updated Review," arXiv:quant-ph/0105101v2 10 Jun 2007. Also see: "Complete description of a quantum system at a given time," J. Phys. A 24, 2315 (1991), and "Time symmetry in the quantum process of measurement," Phys. Rev 134, B1410 (1964).
[35] Y. Aharanov, D. Albert, and L. Vaidman, " How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100," PRL 60 1351-1354 (1988). [AAV]. (alse see Physics Today, Nov. 2010, and Physics World, 3/1/13 p. 40).
[36] Yakir Aharonov, Peter G. Bergmann, Joel L. Lebowitz, "Time symmetry in the quantum process of measurement," Physical Review B., vol. 134, no. 6, pp. 1410-1416, 1964. [ABL].
[37] Ken Wharton, "Quantum Theory without Quantization," arXiv:1106.1254v1 quant-ph 7 Jun 2011.
[38] K.B. Wharton, "Time Symmetric Quantum Mechanics," Foundations of Physics 37, No 1, Jan 2007, p 159-168.
[39] G. Rajasekaran, "Are Neutrinos Majorana Particles?", arXiv:0803.4387v1 [physics.pop-ph] 31 Mar 2008.
[40] "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 .
[41] Alain Aspect, J. Dalibard, G. Roger: Experimental test of Bells inequalities using time-varying analysers. Phys. Rev. Lett. 49, 1804 (1982). Free access: http://www.phys.ens.fr /∼dalibard/publications/Bell_test_1980.pdf
[42] A. Zee, Quantum Field Theory in a nutshell, Princeton University Press, 2003.
[43] Feynman, Leighton, Sands, The Feynman Lectures on Physics, Volume 1, California Institute of Technology, 1963, Chapter 31.
[44] C.D. Froggatt, H.B. Nielsen, "Trying to understand the Standard Model parameters," arXiv:hep-ph/0308144v2 16 Aug 2003.
[45] George F. R. Ellis, "On the limits of quantum theory: contextuality and the quantum-classical cut," August 29, 2011, http://arxiv.org/pdf/1108.5261
[46] Frank Wilczek, The Lightness of Being, Mass, Ether, and the Unification of Forces, Basic Books, 2008.
[47] Frank Close, Neutrino, Oxford University Press, 2010.
[48] David Deutsch, The Beginning of Infinity: Explanations that Transform the World, 2011, Allen Lane.
[49] Steven Weinberg, "Collapse of the State Vector," ArXiv quant-ph 29 Sep 2011.
[50] Brian Whitworth, "A Virtual Reality Conjecture," 2011, http://arxiv.org/pdf/1110.3307
[51] Fotini Markopoulou, "The Computing Spacetime," gr-qc 17 January 2012, http://arxiv.org/ pdf/1201.3398

[52] Anton Zeilinger, "On the Interpretation and Philosophical Foundation of Quantum Mechanics, http://www.quantum.at/fileadmin/zeilinger/philosoph.pdf.
[53] John Baez, "The Octonians," Bulletin Am. Math. Soc., 39, No. 2, p 145+, 21 Dec, 2001.
[54] The Great Courses: Quantum Mechanics, The Physics of the Microscopic World, Benjamin Schumacher, 2009, DVD series.
[55] Brian Greene, The Elegant Universe, superstrings, hidden dimensions, and the quest for the ultimate theory, W.W. Norton & Co., 1999 (e.g., Fig. 8.10, pg. 208).
[56] Gerard 't Hooft, "How a wave function can collapse without violating Schr/"odinger's equation, and how to understand Born's rule," arXiv:1112.1811 [quant=ph] 20 Jul 2012.
[57] J-S. Boisvert and L. Marchildon, "Absorbers in the Transactional Interpretation of Quantum Mechanics," arXiv:1207.5230v1 [quant-ph] 22 Jul 2012.
[58] Sacha Kocsis, et. al., "Observ ing the Trajectories of a Single Photon Using Weak Measurement," 19th Australian AIP Congress, 2010 (also see Science 332 no 6034 pp1170-1193, June 2011).
[59] Don Howard, "Albert Einstein, Physicist, Philosopher, Humanitarian," The Great Courses, Lectures 1-24, 2008, DVDs.
[60] E. Megidish, et. al., "Entanglement Between Photons that have Never Coexisted," arXiv:1209.4191 [quant-ph] 19 Sep 2012.
[61] Andrei Khrennikov, Växjö Interpretation of Wave Function: 2012, arXiv:1210.2390 [quant-ph] 7 Oct 2012
[62] Roger Bach, et. al., "Controlled double-slit electron diffraction," arXiv:1210.6243v1 [quant-ph] 22 Oct 2012.
[63] Michael B. Heaney, "A Symmetrical Interpretation of the Klein-Gordon Equation," arXiv:1211.4645v1 quant-ph 20 Nov 2012.
[64] M. Schlosshauer, J. Kofler, A. Zeilinger, "A Snapshot of Foundational Attitudes Toward Quantum Mechanics," ArXiv:1301.1069v1 [quant-ph] 6 Jan 2013.
[65] George F R Ellis, "The arrow of time and the nature of spacetime," arXiv:1302.7291v1 [gr-qc] 28 Feb 2013.
[66] Robert B. Griffiths, "A Consistent Quantum Ontology," arXiv.org 1105.3932 [quant-ph] 12 Feb 2013.
[67] B. J. Hiley, "Bohmian Non-commutative Dynamics: History and New Developments, ArXiv:1303.6057 [quant-ph] 25 Mar 2013.
[68] J.F. Roch, A. Aspect, et. al., "Experimental realization of Wheelers delayed-choice Gedanken Experiment," ArXiv:quant-ph/061023 28 Oct 2006, or Science 315:966-968.
[69] Travis Norsen, Sarah Nelson, "Yet another snapshot of foundational attitudes toward QM," ArXiv 1306.4646 quant-ph 19 Jun 2013 (for 76 contributors to a Quantum Theory without Observers III conference in Bielefeld, Germany, recent poll number 3).

DEFINITIONS:

Entanglement: The usual definition of entanglement is as follows: Quantum entanglement, also called the quantum non-local connection, is a property of the quantum mechanical state of a system containing two or more objects, where the objects that make up the system are linked in a way such that one cannot adequately describe the quantum state of a constituent of the system without full mention of its counterparts, even if the individual objects are spatially separated. This interconnection leads to non-classical correlations between observable physical properties of remote systems, often referred to as nonlocal correlations. Einstein famously derided entanglement as "spukhafte Fernwirkung" or "spooky action at a distance". Schrödingers name was Verschrankung or cross-linking (or 'shared enclosure'). Entanglement is a key trait demonstrating an entire departure from classical physics. John Cramer treats entanglement as communication back and forth in time between detected particles and their source. My preference for the entanglement of particles is that of sharing a list of possibilities in superposition subject to a joint conservation law and that they have a common overlap or set of overlap points such that they can be connected via a time zig-zag pathway (a 'Cramer-path'). There are even recent examples of 4-photon trajectories where two photons become entangled such that the remaining photons do not share coexistence at the same time. One photon is created and measured prior to the other being created and measured still giving full quantum correlations [60]. Without backwards and zig-zag communication in time, this would be truly weird.

By definition, an entangled state cannot be written as a tensor product of states. [20] Quantum mechanics allows multiple particles or qubits to be in quantum superposition which for the case of n = 2 may resemble a "Bell state combination like $\psi = 00 + 11$, $00 - 11$, $01 + 10$, or $01 - 10$. Maximally entangled two-photon states are also called Bell states. As a short distance example, the two electrons of a helium atom are entangled (measurements cannot be made on one particle without affecting the other). And the two atoms of a hydrogen molecule are entangled. Presently, longer distance separations are of more interest. It is believed that the two gamma rays from the decay of positronium will be entangled, and the two bottom quarks from the decay of the upsilon particle are entangled (Belle, 10 GeV, 2007). For single particles, different degrees of freedom can be entangled: polarization and path for single photons, spin-path-energy for single neutrons in a perfect Si-crystal interferometer (2012). Quantum mechanics violates Bell's inequality; but separable states do not violate any Bell's inequality. Einstein's primary concern was the "principle of separability [59]," but quantum mechanics can indeed have the non-separability of entanglement.

Entanglement has been experimentally verified many times now for the case of photon polarizations (e.g., Clauser-Freedman 1972, Alain Aspect 1982 − using photon cascades from Ca and Hg atoms). Today, the most frequent technique for generating entangled photon pairs is parametric down conversion (PDC). A new technique is entangled photons on-demand from biexciton cascade of single quantum dots. Quantum entanglement has been achieved between single atoms separated by a distance of a meter (Christopher Monroe, 2007). Confined ions can be entangled (e.g., Yb, 2007), and NIST has entangled atom pairs (Be, 2006). In 2011, 14 quantum bits were entangled as a sequence of calcium atoms in an ion trap and manipulated using laser light. Entanglement in superconducting circuits has been limited so far to two qubits, but up to ten qubits for photons. Entanglement does not occur in liquid NMR. And it appears that entanglement has not yet been verified for the very important case of electron spins (although that may happen soon). Rather than just being a problem to be understood, entanglement is a requirement for universal

[20] A Pure State psi is entangled if $|\psi\rangle \neq |\psi_1\rangle \otimes ... \otimes |\psi_N\rangle$. For example, $GHZ = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\uparrow\rangle + |\downarrow\downarrow\downarrow\rangle)$ or $W = \frac{1}{\sqrt{3}}(|\uparrow\uparrow\downarrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\downarrow\uparrow\uparrow\rangle)$.

quantum computing.

Weak Measurement: DEF: A 'Weak Measurement' is a determination of the mean value of an observable which is only weakly interacting with a system. It is found by the equation, $A_w = \langle f|A|i\rangle / \langle f|i\rangle$, where $|i\rangle$ is the pre-selected initial state preparation, $\langle f|$ is the post-selected final state, and A is the weak observable. Single measurements for that observable then have very low precision and wide variation, but a large sub-ensemble of such measurements can have a well-defined average value. An example of a weak measurement for electron spin is a Stern-Gerlach weak magnetic field gradient applied in a z-direction while the final measurement is strong in the x-direction. The result will show a small measureable z-deflection of the beam.

Notice that the pointer shifts predicted for weak measurements are inversely proportional to $\langle f|i\rangle$, the overlap of the initial and final states. So if the overlap is small, the pointer shift may be extremely large larger than could ever occur without post-selection. This is called weak value amplification, but it also reduces the frequency of occurrence of the post selected events. So, high statistics are required by repeating a process many times on identically prepared system. Without post-selection, weak measurement must agrees with the standard quantum formalism. For interpretation, one is tempted to say that the mean value represents something really out there (such as an average velocity).

The concept of weak measurement in 1988 evolved from a previous work by Aharonov on the "Two-State Vector Formalism" [TSVF, 1964] [35][36] This is a time-symmetric formalism of quantum mechanics where $\langle f|$ is a backwards in time evolving quantum state. An in-between 'now' state depends both on the past and also on measurements made in the future. This time symmetry has similarities with the transactional interpretation of John Cramer (1986) and could represent physics after a transactional selection.

Emergence: Professor Jeffrey Goldstein in the School of Business at Adelphi University provides a current definition of emergence as: "the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems" (Corning 2002). Goldstein's definition can be further elaborated to describe the qualities of this definition in more detail: "The common characteristics are: (1) radical novelty (features not previously observed in systems); (2) coherence or correlation (meaning integrated wholes that maintain themselves over some period of time); (3) A global or macro "level" (i.e. there is some property of "wholeness"); (4) it is the product of a dynamical process (it evolves); and (5) it is "ostensive" (it can be perceived). For good measure, Goldstein throws in supervenience – downward causation." A discussion of the emergence of higher levels of reality from lower levels is given by George Ellis [45].

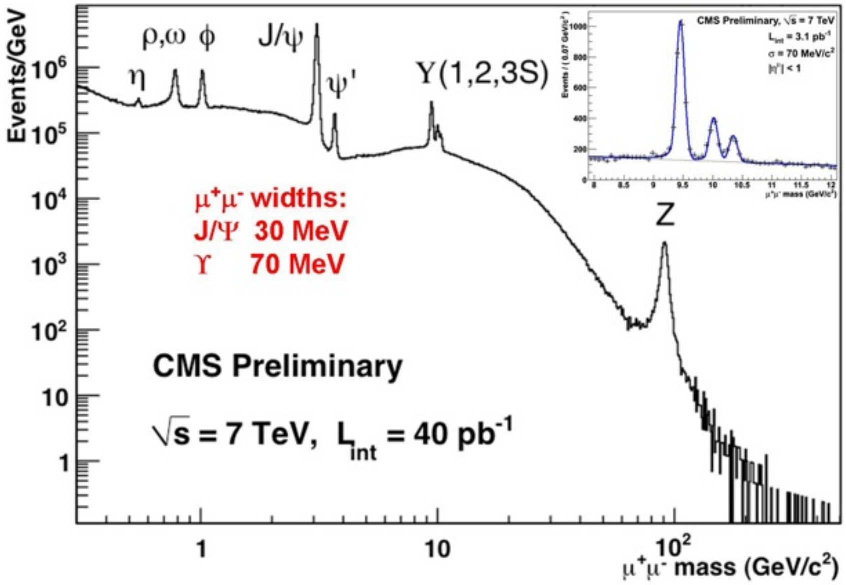


FIGURE 1. LHC-CMS DiMuon output progression of different resonance states, T.Dorigo, 24 Jan 2011

10. "Here there be Dragons"

One of the most familiar quotations about the dangers of trying to interpret quantum mechanics was given by Robert Griffiths [66] who wrote:

"One can be sympathetic with strict orthodoxy in that it is intended to keep the unwary out of trouble. Careless thinkers who dare open the black box will fall into the quantum foundations swamp, where they risk being consumed by the **Great Smoky Dragon**, driven insane by the Paradoxes, or allured by the siren call of Passion at a Distance into subservience to Nonlocal Influences. Young scientists and philosophers who do not heed the admonitions of their elders will, like the children in one of Grimms fairy tales, have to learn the truth by bitter experience."

He adds that "interpretation of quantum theory found in textbooks is widely regarded as quite unsatisfactory. Among philosophers of science this opinion is almost universal, and among practicing physicists it is widespread. It is but a slight exaggeration to say that the only physicists who are content with quantum theory as found in current textbooks are those who have never given the matter much thought, or at least have never had to teach the introductory course to questioning students who have not yet learned to 'shut up and calculate!' "

Similarly, Steven Weinberg is dissatisfied with Copenhagen and also with Many Worlds (which has no explanation of the Born Rule). Regarding interpretations in general, "It would be disappointing if we had to give up the "realist" goal of finding complete descriptions of physical systems, and of using this description to derive the Born rule, rather than just assuming it." It is hard to live with no description of physical states at all, only an algorithm for calculating probabilities. "Today there is no interpretation of quantum mechanics that does not have serious flaws." There should be greater awareness of this greatest of all unresolved problems in science. It is a shame that very few working physicists, even theoreticians, have thought carefully and deeply about quantum foundations.

11. Appendix: Collapse Analogies:

It is possible that collapse entails actual use of random numbers by Nature. But many physicists believe that collapse is an emergent result from a complex system. In case an actual physical transport of particle properties is really required, consider the following: Between an emitter and absorber, it is possible that sub-physics actual path exists for a quantum transfer and involves something akin to a break-through phenomenon in space-time. It may be that the terms: wave-function collapse, emergence, symmetry breaking, butterfly-effect, breakthrough, paths-convergence, selection, and environment-induced-superselection or environmental-selection ("einselection" [29]), have some overlap to approximately mean the same thing for this concept. It might be profitable to consider individual space-time network pathways and their information contents as agents in a complex-adaptive-system (or CAS) so that breakthrough transfer is an emergent phenomenon. There are some familiar analogies from common experience which shed a little light on the concept of path convergence: foraging ants, foraging bees, and lightning leader strokes.

In the **foraging ant analogy**, ants are agents in a complex adaptive system whose scientific literature mentions topics like decentralized communication, trail connectivity and emergent benefits of ant pheromone trail networks. Foraging individual ants deposit chemical messages on the ground so that other ants can follow previous trails towards sources of food. When a forager succeeds in finding food and returns, it touches antennae with other ants to encourage them to also go out for the food. Collectively this facilitates self-organization with additional ants leaving reinforcing trails which tend to be straighter. Positive feedback then results in the emergence of a heavily traveled path between the colonys nest and a food source. Even in the case of a symmetric circle of sugar about an ant-nest hub, only a few strongly used final pathways will emerge [20]. The pheromone trail is a temporary environmental memory which gradually evaporates and weakens. One basic term in the rich field of studies of social insects like ants is "stigmergy" which refers to the collection of mechanisms in which individual entities socially influence the actions of others by using indirect communication aided by subtly modifying their environment. And a frequently mentioned example of "stigmergic communication" is the pheromone trail laying behaviour of ants. Collectively, a colony emerges novel patterns and structures evolving an overall coherence well beyond that encoded in the individual.

The **communication of bees** may be more intricate than for ants because they can't lay trails on the ground. Scout bees fly through the air in search for food and then transmit their discoveries to the hive by an elaborate figure eight style dance with different levels of enthusiasm and endurance over the perceived value of the food to the colony. Worker bees gather around in groups of perhaps five individuals and pay attention to the foragers and their competing messages. The insistence of the dancer can be contageous and can inspire other foragers to fly out and investigate. They too may come back and reinforce the message with their informative dances which include information on distance and relative location to the food sources. One of the competing bees wins the collective decision of the hive which then goes out as a pack to harvest a flower patch [18]. Karl von Frisch won the Nobel Prize in 1973 for deciphering bee communication including the meaning of their "waggle dances."

These are socio-biological examples, but an analogy to **lightning** may be more physically relevant. Lightning uses a bifurcating exploration of pathways for a satisfactory ultimate high-powered path. Also involved is short-term memory of a previously ionized leader path which is available for reuse, refresh, and evolution of subsequently longer pathways. In a lightning storm, there is a need to re-balance electrical charge between a cloud to ground or to another cloud. High energy channels of ionized air get established from source to sink resulting in a path of least resistance for subsequent large current coulomb level transfers. There is then a stroke between earth to cloud leader and then a more luminous return stroke discharge. High speed videos show multiple strokes covering the same established pathway producing a notable strobing effect. Response strokes are preceded by intermediate dart leader strokes similar to but weaker than the initial stepped leader. A downward-propagating leader is always forked with only one fork surviving at each point of the downward meander [19].

These examples show evolution towards selected pathways but suffer from using low populations– of foraging ants, of hive size, of total forking numbers. For relevance to quantum mechanics, much higher numbers of elements or segments and a selection mechanism is required. Also, all of these examples occur in forwards time development. But quantum mechanical path development and choice seems to be nearly instantaneous. One article says that this development must occur in "hidden time" [18]. John Cramer uses a similar term "pseudo-time"for his back and forth transaction communications. Reinforcing a selected pathway requires the existence of some kind of memory in space-time.

There is one interpretation of QM [61] in which prequantum random signals cross a threshold and then cascade into collapse in a chosen detector (TSD = measurement threshold detection model). There exist intrinsic processes in the microworld leading to discretization of continuous background fields with the act of discretization or quantumness being created in detectors through the act of measurement. This interpretation (from

"Växjö" in Sweden) says that a wave function describes correlations in prequantum random fields.

12. NOTES:

A major lesson from "decoherence" studies is that all environments are noisy and that noise can degrade or break quantum superpositions– make them decohere within a short decay time [28]. Entanglement is required to do quantum computing, but decoherence takes it away [31]. Examples of background noise include randomly fluctuating electric fields, phonon vibrations, electron-phonon coupling, microwave background, fluctuations in magnetic exchange fields, AC power fluctuations, voltage gate fluctuations, background zero-point fluctuations, Brownian motion, and frictions and viscosity dissipation. A recent example of Gaussian background is varying current in coils for magnetic fields for quantum registers to study calcium ions in a Paul trap environment. Although Paul traps don't normally use magnetic fields, recent tests at Innsbruck University had a bias magnetic field for lifting degeneracy of ground and excited states. Decreasing this noise strongly improved experimental single-qubit coherence time [30]. Another experiment showed that Buckyball fullerene double slit interference decoheres as gas is introduced into a testing vacuum. Decoherence serves to change coherent alternatives into classical mixtures of alternatives– but that doesn't quite solve the collapse problem into one single classical result − only 15% of physicists said the measurement problem is solved by decoherence.

Note that being in two places at once is just a special case of having a superposition of alternatives −many states at once.

The Vacuum machine that does the sensing of quantum wave-function information also does the broadcasting, distributed networking, constitutes the nodes and links and information processing. It can be activated over a very wide range of 'quantum energies' whose vibrations simulate particles.

As a lesson from Bohm, recall that a Bohm particle in a rectangular potential well has no velocity − it just sits there. If one side of the well is removed, then the quantum potential accelerates the velocity to a value corresponding to classical expectations. So, in the quantum world, one can multiply m x v or m/2 times v^2 − but the products are not necessarily expected "real" p or KE. What we call m, p, E, or charge have classical bias. In the quantum world, they may be something else. For example, quantum reality may be information obeying QM equations, and classical expectations only appear at set-ups and classical measurements. If m, p, E, q, are really quantum reality information attributes, then there is no longer a problem with superpositions. Possible paths interfere. But, an infinite number of possible paths for an electron each having mass and charge do not result in infinite net mass and infinite net charge from real duplications. These things only exist when the possibilities have been resolved to single outcomes.

The bouncing droplets scenario for classical wave-particle duality has the droplets creating all the guide waves. For Bohm, the particles are passive − they create no waves .

Vigier [dp 8/10] suggests that particles beat in phase with their surroundings so that there is an interaction but on the particle. Now, by rotating the vibrating table, analogies to Zeeman splitting can be produced classically (July, 2011).

Double-slit electron diffraction interference patterns were obtained in 2012 using the Feynman example with a movable mask in front of a double slit to control the output probability distribution for single versus double slits [62]. Previous experiments used electron biprisms to split an incoming electron beam, but this experiment used FIB (Focused Ion Beam) milled nano-slits on gold coated silicon-nitride. The slits were 50 nm wide separated center to center by 280 nm. A 5 micron wide mask could be moved into position using a piezoelectric actuator. Single electron build-up patterns were recorded.

A Bohmian solution could mean that an actual path is selected for something resembling an actual particle. But it is a separate rule that only one of several possible Bohmian paths carries a particle. Having it occur in separate paths at the same time smacks of Many Worlds.
Pauli's challenge to de Broglie was about inelastic scattering (1927).

WAVEFUNCTION SUB-QUANTA INFORMATION

DAVE PETERSON

Abstract. It is proposed that the quantum wavefunction is a carrier for quantum information with an emphasis here on processing by Nature more than by a detector/observer. A 'psi-wave' of almost any amplitude is a **code** enabling knowledge of momentum, energy, and represented particle-mass. Information for some other particle properties do not have to be carried in psi or by physical moving particles themselves because knowledge of attributes such as electric charge, e, or "weak charge," or magnetic moment are available everywhere in the underlying relevant quantum fields filling all of spacetime. Beyond that, psi represents field disturbances and "simulations" over possibilities. It is an holistic entity communicating broadly within itself instantaneusly.

1. Introduction

[Niels Bohr] "There is no quantum world. There is only an abstract quantum physical description. It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature [1] [2]*."*

[E. T. Jaynes] "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 omelette that nobody has seen how to unscramble. Yet we think that the unscrambling is a prerequisite for any further advance in basic physical theory. For, if we cannot separate the subjective and objective aspects of the formalism, we cannot know what we are talking about; it is just that simple." 1996, [2].

[Stephen Weinberg] "...today there is no interpretation of quantum mechanics that does not have serious flaws..." [3] *2013.*

The purpose of the present paper is to advance potential understandings of the inner workings of quantum mechanics beyond and below the level of knowledge available in textbooks. We discuss elementary waves, the nature of the Vacuum and its Fields, the Born Rule, wave-function "collapse," the photon wave-function, how quantum field theory helps understanding of quantum mechanics, composite particles acting as single particles, what can lie beneath conventional quantum theory, and possible limitations on the mathematical

Date: 20 November, 2013, paper updated to: 11 April 2014 .
email: davepeterson137@gmail.com.

perfection of standard quantum mechanics.

Standard orthodox quantum mechanics (the old positivistic Copenhagen Interpretation) insisted that a wavefunction $\psi(r,t)$ was a state specification containing all the information that can be known about a system as a result of measurements (called "ψ-complete") [1]. The question, "Whose knowledge?" and "Whose information?" was answered as, "The observer's ." Quantum mechanical calculations using the wavefunction are processed according to the still somewhat mysterious Born Rule to yield statistical probabilities of occurrences. The 'dogma' from Niels Bohr was that we should deal only with what we can measure.

The meaning of the wave-function is central to on-going debate, and philosophers call this debate "ontic" versus "epistemic." An ontic state is "real," and an epistemic state is a state of knowledge. Einstein was a realist but advocated ψ-epistemic because he believed that conventional psi was incomplete and had a hidden reality below it. Newtonian phase space points and the classical EM field are ontic. Statistical phase space probability (Liouville states) are epistemic. Mathematically, quantum hidden-variable models are ψ-ontic as also is the non-local de Broglie-Bohm guiding wave. Does a wavefunction packet pertain to an individual particle? Or does it represent an ensemble of identically prepared particles for experimentation? Is its space-time wavepacket profile intrinsic or does it also depend on the observers knowledge? In current literature, these issues are unresolved, strongly debated, and believed to be experimentally unresolvable by conventional quantum mechanics.

In contradiction to standard Copenhagen interpretation, we assume that there is a real mechanism that makes quantum mechanics work prior to collapse and that there is a world describing the way Nature experiences quantum mechanics. This is a belief beyond and below measurement; and the question is, "how far can we take this belief and still have consistency?" What might quantum mechanics look like to Nature with absorbers but the absence of 'observers?'
Modern physics has many new aspects that appear to be beyond even future testability (holography, strings, Planck sizes, gravitons, multiple-dimensions, Hawking radiation, multiverse...). Is there any reason to believe in them? Some advocate that they can gain general acceptance if there are no plausible alternative theories and a theory possesses unexpected explanatory coherence. Eventually, we hope to do that for quantum mechanics.

We claim here that the wave of the wavefunction actually exists as an information carrier that is real in the sense that it is used by Nature. Inbetween these extremes, Ed Jaynes says that a quantum state is a mixture of the partly real in Nature and partly incomplete

[1] Mathematically, the state of a physical system is defined by a ket in the state space (a linear vector space of square-integrable functions −automatically implying the superposition principle). A measurement result is an eigenvalue of an observable. [Of course, neither amplitudes nor superpositions are directly observable. As exceptions to measurement, they are 'grandfathered in' by being basic postulates].

human knowledge, and a goal is to tease them apart [2]. A new theorem by Pusey ("PBR," 2012, [4]) inclines towards the real and essentially says that "individual quantum systems must know exactly what state they have been prepared in, or the results of measurements on them would lead to results at odds with quantum mechanics.

The key ideas in the following discussions are:
The basic entities are selected quantum fields and waves rather than what we have often called "particles." So, in the sentence, "The particle simultaneously takes many separate paths," particle means "field disturbance."
The various quantum fields in space-time possess the basic knowledge and constants of Nature needed for physical processings.
Many of the quantum fields work together as a team (e.g., EM fields and electron charge play together).
Waves are "real," and Nature processes them using derivatives (local phase comparisons) so that absolute phase generally isn't used.
The wave function for a single photon is electromagnetic, obeys Maxwell's equations, and represents an amplitude for electromagnetic energy.

Larger composite "particles" are wavicles upon wavicles and do not have their own "fundamental" quantum fields but can still behave as if they did (e.g., proton field).
Somewhat mysteriously, they possess a wavelength inversely proportional to their total combined mass-energy.
The psi-field is holistic, and psi-wave propagation is time symmetric producing the appearance of instantaneous internal communication.
Entanglements and contextualities suggest sub-quantal communication forwards and backwards in time between emitters and absorbers.
Cosmic "photons that never end" will never be absorbed but still collectively contribute to gravitation. This is a major problem for interpretation and suggests a change needed for the transactional interpretation.

There are some relevant basics that must remain untouched at present: the mechanism for the selection of quantum field collapse (perhaps pure randomness), the detailed values of field quantization (particle masses, energy quantization), and the mechanism for the determination of the arrow of classical time flow (perhaps the thermodynamic limit). These are givens. But does the Born Rule really have to just be a postulate? Some believe it is fundamental and cannot be derived from anything simpler. And there are no convincing derivations at present. There is a challenge about the degree of reality of most of the cosmic photons (CMB, starlight) that have so far traveled freely for billions of years and most likely will <u>never</u> terminate on an absorber. One aspect of their reality is that they do gravitate (contribute to the energy-momentum of general relativity). We suggest that the

[2] But his solution is Bayesian inference where probability is not frequency of random variables but rather logical reasoning in situations of incomplete human information.

"photons that never end" are equivalent to the "photons that have traveled far."

We begin our analysis by first examining the basic case of plane waves:

2. 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 terms a measurement observation might prefer.

$$\psi(x,t) = Ae^{-i(\omega t - kx)} \;\rightarrow\; \psi(x,t) = Ae^{-i(E_o t - p_o x)/\hbar} \tag{1}$$

Since $E = \hbar\omega = h\nu$, and $p = h/\lambda = \hbar k$ (de Broglie's relation for momentum), these equations are equivalent. But suppose 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. That is, the simple wave is everywhere a carrier of 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. The 'particle' itself is only a deduction by the measuring apparatus and 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.

In the wave form with E and p, we can easily state operators that can pull out these observables from the exponent. That is, setting $\hat{p} = -i\hbar\nabla$ gives us $\hat{p}\psi = p_o\psi$. And using $\hat{E} = (i\hbar)\partial/\partial t$ gives us $\hat{E}\psi = E_o\psi$ [3]. And then, of course, expressing conservation of energy for a single particle gives us **Schrödinger's equation** (SE):

$$E = KE + V = \frac{p^2}{2m} + V \;\rightarrow\; -\frac{\hbar^2}{2m}\nabla^2\psi(x,t) + V(x,t)\psi(x,t) = i\hbar\frac{\partial\psi(x,t)}{\partial t} \tag{2}$$

[It has always seemed to me that this simple approach is the best way to intuitively introduce the Schrödinger equation for the first time]. What the $\hat{p}$ operator does is look at the density of wave peaks in space, and the $\hat{E}$ operator looks 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. And we show below that an uncertainty principle applies to waves in general, and that rest mass can also be deduced.

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

[3] And for angular momentum, L, one considers change of phase around a circular phi direction, $\partial/\partial\phi$.

as:

$$(3)\quad \hat{H}\psi(x,t) = i\hbar\frac{\partial}{\partial t}\psi(x,t), \quad or \quad \frac{\partial\psi}{\partial t} = \frac{-i}{\hbar}\hat{H}\psi \quad so, \quad \psi = \psi_o e^{-iHt/\hbar} = U(t)\psi_o(x,0)$$

Where U(t) is a unitary time evolution operator, and we can use the Hamiltonian to give the time evolution of the wavefunction, $\psi(x,t)$.

More traditionally, the Schrödinger equation is simply given as a founding postulate of non-relativistic quantum mechanics. Its solutions include tunneling, entanglements, complex atoms, and s-orbitals which no longer resemble anything like plane waves. For example, just try a solution resembling an exponentially decaying profile: $\psi_1 = Ae^{-br}$ and plug that into the SE with a atomic central potential $V = -Ze^2/4\pi\epsilon_o r$ and $\nabla^2\psi = r^{-2}\partial/\partial r(r^2\partial\psi/\partial r)$. The result is the normalized 1S atomic orbital:

$$(4)\qquad \psi_1(r) = \frac{1}{\sqrt{\pi}}\left(\frac{Z}{a_o}\right)^{3/2} e^{-Zr/a_o},$$

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. 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.

Uncertainty Principle: 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 x versus k (i.e., quantum mechanics not required) or x versus p:

That is, somehow, Nature effectively performs 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 probabililty is described by:

$$(5)\qquad P_x \propto e^{-x^2/2\sigma_x{}^2},\ so,\ \psi(x) = \sqrt{P_x} \propto e^{-x^2/4\sigma_x{}^2}$$

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:

$$(6)\qquad \frac{k^2}{4(1/4\sigma_x{}^2)} = k^2\sigma_x{}^2 = \frac{k^2}{4\sigma_k{}^2},\ \Rightarrow\ \sigma_x\sigma_k = \frac{1}{2},\ \Rightarrow\ \sigma_x\sigma_p = \frac{\hbar}{2},\ \ or,\ \Delta x\Delta p = \frac{\hbar}{2}.$$

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 we have, $\Delta x \Delta k > 1/2$. 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 $\Delta t \Delta \omega \geq 1/2$.

Rest Mass: 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 as dispersion relation. Consider the probably more applicable special relativistic case first:

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:

(7)
$$E^2 = (\gamma m_o c^2)^2 = \frac{(m_o c^2)^2}{1 - v^2/c^2} = (m_o c^2)^2 \left[1 + \frac{v^2/c^2}{1 - v^2/c^2}\right] = (m_o c^2)^2 + (\gamma m_o v c)^2 = (m_o c^2)^2 + (pc)^2.$$

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:

$$\nu^2 = (\gamma \nu_o)^2 = {\nu_o}^2 + \nu^2 c^2 \frac{v^2}{c^4} = {\nu_o}^2 + \left(\frac{c}{\lambda}\right)^2, \ or \ \ \omega^2 = {\omega_o}^2 + (kc)^2. \tag{8}$$

This can be conviently pictured by right triangles with hypotenuse E and sides $(m_o c^2)$ and (pc) (or ν with sides ν_o and (c/λ)) [4].

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.

The Schrödinger equation is non-relativistic with $KE = p^2/2m$. In that case, angular frequency would be written as $\omega(k) = (\hbar k/2m)^2 + V(x)/\hbar$. Then group velocity is $v_g = v = \partial\omega/\partial k = \hbar k/m$, and phase velocity $v_\phi = \omega/k = \hbar k/2m$. Then $v_\phi = v/2$ (which seems very non-physical), and $\nu\lambda = v/2$. $p = h/\lambda = mv = mv_g$, so mass $m = h/v_g\lambda$. And

[4]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 two forms (eqns. (7) (8)) are covered by one 4-vector equation, $p^{\,\mu} = \hbar k^{\mu}$.

The Compton Effect as just waves without particles: The Compton Effect of 1923 for x-rays scattered from essentially free electrons in atoms proposed that the photons act like particles colliding like billiard balls with particulate electrons in atoms. It used Einstein's ideas of light corpuscles of energy $E = h\nu$ and momentum $p = h\nu/c$ and relativistic electron energy of $E = \sqrt{{m_o}^2c^4 + p^2c^2}$ to derive a scattered photon 'final-minus-incoming' wavelength shift $\lambda' - \lambda = \Delta\lambda = (1 - \cos\phi)\lambda_c$ for an angle of photon scattering, ϕ. The Compton wavelength of the electron is defined as $\lambda_c = h/m_oc \sim 2.426$ pm. Most introductory textbooks say that this effect along with the photoelectric effect validate Einstein's photon corpuscle hypothesis.

But, just the next year, in 1924, de Broglie put forth his idea of "matter waves" as distinct from corpuscles. The Compton effect can be derived solely in terms of quantum waves without introducing particles at all [6].

That is, instead of initial and final particle energy and momentum, we consider a plane electromagnetic x-ray wave incident from the left with wavelength λ encountering the Compton wavelength and frequency of a medium (e.g., graphite) of nearly free electrons. The final scattered x-ray direction with wavelength λ' can be pictured as tilted up to the right by angle ϕ and an "electron wave-packet" velocity $\vec{v_e}$ tilted down at angle θ. The dispersion relation of equation (8) gives $\omega = \omega(k, \omega_o)$ from which one calculate a scattered "wave-packet" speed $v = v_g = \partial\omega/\partial k = kc^2/\omega$. In the absence of potentials, V, this is the same as saying that $v_g v_\phi = c^2$ for matter waves as well as light. The initial and final waves can be written in exponential form as in equation (1); and the details of the wave calculations and discussions can be found in reference [6]. Conservation of energy and momentum are not explicitly needed but rather follow from the scattering and the dispersion relation presumably built into the space-time fields.

3. The Born Rule:

A question from the introduction was, "does the Born Rule really have to just be a postulate?"

Quantum mechanical amplitudes are complex numbers, but classical reality generally uses real numbers. Getting real values from a wave-function can be done by calculating $|\psi|$ or by $|\psi|^2 = \psi^*\psi$ [5] , and Max Born selected the latter in a paper written in 1926 describing a scattering problem. If one intuitively envisions a photon wave as an electromagnetic field, then its energy content should go as $\mathcal{E} \propto E^2$. Since energy [or energy-momentum $p^{\,\mu} = (E/c,\ \vec{p}) = \hbar(\omega/c,\ \vec{k}) = \hbar\, k^{\mu}$] is quantized , all of the energy has to be dumped

[5] In more detail, if an operator A has eigenvalues λ_i and eigenvectors $|a_i\rangle$, so that $A|a_i\rangle = \lambda_i|a_i\rangle$, then the probability of finding $\lambda_i = |\langle a_i|\psi\rangle|^2 = |a_i|^2$. Or, as transition probability from state 1 to 2, $P(1,2) = |(1,2)|^2$. Pragmatically, measurements occur in the lab and probability is given by the frequency interpretation (although that is "held in rather low regard in the philosophy of probability [7]).

(collapsed) into a detector in proportion to its local energy density. In a handwaving sense, this justifies using $P \propto |\psi|^2$. This may be one reason the otherwise mysterious Born Postulate caught on so quickly. The electric field, E, could be considered as an EM "energy amplitude;" it is after a probability amplitude in the same sense as the matter wave ψ, and what is the reality of that? But, It may be that this E field has a different "reality" from the classical E field.

The electron wave-function represents a matter wave that also has to obey the photon quantization condition, $p^{\mu} = \hbar k^{\mu}$. A difference is that an electron also possesses a particular constant value of rest mass or rest frequency as encoded in the electron fermion field. Unlike light, the electron wave $\psi(x, t)$ can be considered as a space-time solution to the Schrödinger equation. The term "probability amplitude" is colored by a view from the collapsed result which is beyond the range of the continuous-and-unitary Schrödinger wave. Perhaps the terms energy-amplitude and matter-amplitude have more physical insight than the vague term probability amplitude.

In general, there is no convincing derivation of the Born Rule, and most attempts may be circular [7] [6].

The existence of entanglements, contextualities, [7] and apparent collapse encourages a consideration of a presently minor interpretation of quantum mechanics called the Cramer **Transactional Interpretation** of quantum mechanics from 1986 ("TI" or "TI-QM") [9]. TI is discussed primarily for photons (supposedly the easiest example) and liberally refers to the symbol ψ. A photon is its own antiparticle and can therefore travel both forward and backward in time in the Feynman sense. It states that an initial **offer wave**, ψ travels from an emitter to a detector, and then a response wave ψ^* travels backwards in time to the source resulting in a "hand-shaking" agreement. And finally, a more localized 'particle' travels forward in time to the detector. An advantage of this latter view is that it seems to be the only scenario that "explains" the Born rule, $\psi^*\psi$ and has any time slice showing both a wave and a particle simultaneously. Also, these transactions are intrinsically non-local and effectively a-temporal.
As an example, an oscillating electric dipole ($p = qd$) in the radiation zone ($d \ll \lambda \ll r$) produces an outward propagating and weakening electric field $E \propto \sin\theta \, e^{i(kr-\omega t)}/r$. Upon trial absorption or scattering, another wave like this is sent backwards in time from the absorber to the emitter so that the net signal at the emitter is detected as $\psi^*\psi \propto (\sin\theta/r)^2$.

[6] There is a largely accepted quantum logic derivation of the Born Rule called Gleason's Theorem from 1957 [8]. But there are also claims that it is not rigorous, quite difficult, has restrictive assumptions, is not insightful, probably circular, and not useful for the Everett interpretation because MWI lacks projections.

[7] John Bell theorem of 1964 and the Kochen-Specker (KS) theorem of 1967. Bell's theorem says that "No physical theory of local hidden variables can ever reproduce all of the predictions of quantum mechanics." And the KS theorem says that two of the basic assumptions of hidden variable for quantum mechanics yield a contradiction: that hidden variables of observables possess definite values at any given time, and that they have values that are intrinsic and independent of measuring devices (non-contextuality).

This is the same form as outward energy density at the detector, i.e., a Poynting vector $S \propto (\sin\theta/r)^2$.

Regarding Jaynes' quotation in the introduction, how can the Bohr-Heisenberg omelette be unscrambled? It is important to first realize that knowledge of the measuring device really is a crucial part of quantum mechanics. What kind of measurement is to be done affects the results of the measurement. A mechanism is needed to explain this, and quantum communication back and forth in time would do it. The formation of a Cramer transaction between a source and an absorber is a key ingredient of the sub-quantum world. The nature of the source and the absorber are both required for what may occur between them. Some "scrambling" is intrinsically needed.

TI is patterned after the Wheeler-Feynman absorber theory for electromagnetic radiation which uses half-advanced (backwards in time) and half-retarded (forward in time) waves. So in TI, a source emits both a ψ-wave forward in time along with a ψ^* wave backwards in time. The receiver also does this to provide both $\psi*$ response waves and new ψ retarded waves too. But the phase-shift of the retarded wave from the receiver cancels the sender's retarded wave beyond the absorber. And the advanced waves before the emitter are similarly canceled by destructive interference. The unwanted extraneous signal waves are canceled out leaving only waves between the emitter and absorber. That is, Cramer wished to address the mysterious "collapse" phenomenon mechanism by nullifying unneeded portions of the initial offer wave so that their information could not be reused. But in retrospect, but that was perhaps too ambitious. Another possibility is the sub-quantal holistic web cancels the wide ranged wave and arranges energy transfer to a point. Another possibility is that the first offer wave is canceled or ignored while a new wave is then separately broadcast which is more directed and localized. The first wave is just for information gathering, while the second wave is for "real."

Our understanding of the universe has evolved unexpectedly since the TI paper from 1986. We now have accelerating expansion and new ingredients of dark energy and dark matter which does not interact electromagnetically. That means that a full theory depending on complete absorption of offer waves like that of the Wheeler-Feynman absorber theory is unlikely to be correct. It now seems that there are **'photons that never end'**: most of the free propagating photons in the ever expanding universe will never be absorbed (e.g., CMB photons and starlight). These waves have been traveling through the universe for many billions of years, and the future universe is increasingly sparse. Strangely, no one seems to have considered this obstacle for TI. But the initial idea that psi-waves can go both forward and backward in time and facilitate a communication is still very attractive and may still be maintained over distances that are *not too long*. And this naturally leads to the Born probability from a complete hand-shaking transaction. All of these photons (radiation in general) still contribute to the Einstein field equations (they gravitate). Even though they never encounter an absorber, they still have a measure of existence. That poses a puzzle for our understanding of quantum mechanics. The question of the degree of "reality" of these photons is complicated by disagreements about conservation of energy for this radiation. Has the cosmological redshifting of the radiation gone into the background gravitational field? Is this shift an observational effect? Or is energy actually lost? Note that there is no such thing as the density of gravitational energy. In addition, some say "reality is unborn until is is observed [38]." That goes with Asher Peres', "unperformed experiments have no results." Perhaps that means that various possibilities have not become concrete, and one cannot safely deduce counterfactuals. But certainly some degree of reality remains. In the last section of this paper, we suggest that "photons that never end" have the same degree of reality as "photons that have traveled far." In addition to detected energy, the polarization of CMB photons has also recently been detected by researchers at the South Pole (both E-mode and now B-mode as well, [39][40]).

4. Collapse:

Prior to collapse or reduction of a wave function, the choice of a particular absorber out of many possibilities has to be made. That selection could indeed be purely and intrinsically random as per the standard dogma or it might be emergent, but no one is presently prepared to explain the underlying physics of that. Some say that decoherence theory explains it, but this still uses entanglements which in turn request an explanation mechanism. It is said that "the decoherence approach has not solved the measurement problem" [10] and no unitary treatment of the time dependence can explain why only one dynamically independent component is selected (and experienced).

Supposing that selection is somehow accomplished, how can collapse then be explained? We already have the phenomenon of entanglement as a partial analogy to collapse. A measurement on one of a pair of entangled particles instantaneously alters the detection results of the other particle. The most intuitive explanation for this is propagation of sub-quantum information back and forth in time along both paths of the pair of particles connected to a source (a 'sub-quantum-web' view). That propagation might occur at the speed of light ($\pm c$) and still give the appearance of instantaneous communication. Einstein called this "spooky action-at-a-distance." Why not carry this property back from several particles to the parts of just one single offer wave so they are all one single interconnected holistic correlated entity. All space-time regions of an offer wave are in intimate ongoing contact with all other regions. The selection of a single ray from a source to an absorber can then instantaneously result in the cancellation of the other rays or portions of the offer wave by the internal programming of the web rather than by phase cancellation. The information for the transmission of a single quanta can only be used once and then it is gone. Now, we are used to saying that all parts of a propagating EM wave possesses energy, and we consider energy to be the fundamental measure of existence. The idea that widely diffuse energy can suddenly vanish and appear as one quanta of energy is difficult, but energy transfer to a detection point is an experimental fact. The mechanisms of Nature could do that using its apparent action-at-a-distance collapse, and conservation of energy is really supported at the end points of a transaction. What takes place in the middle could be quite non-intuitive.

Roger Penrose (2005) has this to say about ψ-waves: "Wavefunctions are quite unlike the waves of classical physics in this important respect. The different parts of the wave cannot be thought of as local disturbances, each carrying on independently of what is happening in a remote region. Wavefunctions have a strongly non-local character; in this sense they are completely holistic entities [11]."

If sub-quantum mechanics explores or simulates all future possibilities, perhaps it also explores many opportunities for Collapse as well prior to the materialization of a full quanta. This is similar to casing out various banks before you actually rob one. Perhaps collapse is a basic construct so that Nature is often attempting "spontaneous localizations. These fluctuations themselves might not have to be real but rather just simulation attempts − so they may not contribute to the cosmological constant Λ. If collapses occur once at a detector, they may just be a natural propensity that could happen with matter of any size. There are interpretations of quantum mechanics dealing with these spontaneous collapses; and sometimes they involve adding new nonlinear and stochastic terms to the quantum evolution equation [10]. Since they are non-standard modifications, they may ultimately be testable.

5. A Wave Function for the Photon

This is a very basic topic that doesn't seen to be addressed by any text on quantum mechanics but really should be. Quantum mechanics began with the analysis of photon black body radiation and its introduction of the Planck constant, h. Many tests of the foundations of quantum mechanics are performed with photons. And many texts use photons as examples for basic quantum mechanics without going into details. And, for practical purposes, it <u>is</u> possible to write out special wavefunctions that work for photons as long as certain considerations are stated such as locking in a constant number of photons at one or two, having reasonably low energy, and having the photon field only be a transverse vector field.

The wave function for a single photon (PWF) is a different entity from a matter wave function because: a photon has no mass, it is never non-relativistic, photon number conservation isn't required, it has spin one, it is its own antiparticle, and there is no position operator for the photon [8]. Its field parts travel at one speed, c; so there may be no wave packet spreading like that for matter waves. Photons are freely emitted and absorbed, and that topic is covered by quantum electrodynamics (QED) second quantization creation and

[8]The issue of localization is difficult in quantum field theory both for free fields and also for relativistic electrons so that "the electron is as bad as the photon" [Peierls, 1973]. But localization is a key attribute of particleness.

annihilation operators. Because of all this, some claim that there is no such thing as a photon wave function in space-time.

But, there are a variety of ways of stating useful photon wave functions. The best developed system is from the works of Bialynicki-Birula [35] or Raymer and Smith [36]. Following the Dirac ideas but for spin-one, they begin with the Einstein equation to Klein-Gordon operator equation to linearized versions for the case of $m = 0$.

$$E = \sqrt{(mc^2)^2 + (cp)^2} \qquad \rightarrow \qquad E\psi = i\hbar\partial\psi/\partial t = c\sqrt{\vec{p}\cdot\vec{p}}\,\psi.$$

If momentum is considered as an operator, there are two ways to take the above operator square root to give the following equation:

$$i\hbar\frac{\partial\psi}{\partial t} = c\hbar\nabla\times\psi = -i\hbar c(S\cdot\nabla)\psi(x,y,z,t) \tag{9}$$

This assumes that $\psi = \psi_L + \psi_T$ is only transverse ($\vec{p}\cdot\psi_T = 0$ and $\vec{p}\times\psi_L = 0$) and that $S = (s_x, s_y, x_z)$ are spin-one 3×3 matrices [for example, the three anti-symmetric out of 8 total generators of SU(3)].

So, what could ψ be? We know that a single photon can travel and bend through an assembly of glass components just as would a classical light ray. That means that the photon acts <u>as if</u> it were an electrical wave. What is waving is the quantum vector potential of the photon-field or the quantum field $\vec{E} = -\partial\vec{A}/\partial t$. This processing (e.g., determining the index of refraction through glass and making use of it) is performed by machinery below the level of observability. Try the following six-component "Riemann-Silberstein" (1907) electromagnetic field form for the photon wavefunction and then substitute it into the curl equation above with the following result:

$$\psi(x,y,z,t) = \sqrt{\frac{\epsilon_o}{2}}(\vec{E}\pm ic\vec{B}),\ \Rightarrow \nabla\times\vec{E} = -\frac{\partial\vec{B}}{\partial t},\ and\ \nabla\times\vec{B} = \frac{\partial\vec{E}}{c^2\,\partial t}. \tag{10}$$

But we recognize these equations as the free space Faraday and Ampere Maxwell equations. So the photon wave equation can yield the source-free equations for electromagnetism [9]. In addition, energy density is given by:

$$\mathcal{E} = \tfrac{1}{2}(\vec{E}\cdot\vec{D} + \vec{H}\cdot\vec{B}) = \tfrac{1}{2}(\epsilon_o E^2 + B^2/\mu_o) = \tfrac{\epsilon_o}{2}(E^2 + c^2B^2) = \psi^*\psi = |\psi|^2. \tag{11}$$

So, the solution ψ can be interpreted as an **amplitude for the electromagnetic energy density**. The Born rule is energy density which somehow becomes probability. It is not the so-called "photon-particle" that is localized during travel, it is a spread-out energy amplitude that lives in space-time coordinates. And it would appear that Maxwells theory of electromagnetism is the relativistic quantum theory of a single photon in disguise. For simple light interference effects, a photon can be approximated as an E-field wave.

[9] the $\pm$ refer to the two helicity states of the photon. It can also be shown that $\nabla\cdot\psi = 0$ so that $\nabla\cdot\vec{E} = 0$ and $\nabla\cdot\vec{B} = 0$ as well.

If we picture a "photon" as a time-window "wave-packet," it is not like the dispersive Schrödinger matter wave-packet with a faster leading edge and slower trailing edge. In free-space, the photon or EM wave travels only at the speed of light, c. As far as we can tell, the pulse shape would be preserved over long travel times. There are many examples showing that massless radiation has very limited dispersion over astronomical distances. As two basic cases: there are short Gamma Ray Bursts (GRBs) from neutron star mergers billions of light years away having short bursts within a narrow window of typically 0.2 seconds. That certainly limits the degree to which the fields can spread out in space and time. And also, on 24 February, 1987, 11 supernova anti-neutrinos were detected at Kamiokande II in a short burst of 13 seconds. At the same time, there were 8 seen at IMB and 5 seen at Baksan representing a core-collapse for SN 1987 A in the Large Magellanic Cloud dwarf galaxy. So even for slight masses, there is little spreading out of any neutrino "wave-packets."

There are other representations than the 3×3 matrices just discussed. In 1939, Nicholas Kemmer [42] wrote equations for spin-0 (5×5 matrices) and spin-1particles (**10×10** matrices). With a little care, these equations still work up to the present day. Improved equations (Duffin-Kemmer-Petiau or "DKP") are relativistic first-order equations motivated by the form of the Dirac equation and Proca's equation and for spin-0 has equivalence to the Klein-Gordan equation (proven for S-matrix calculations including electromagnetic fields and spin-0 meson calculations including one-loop corrections). The spin-1 equation with 10×10 matrices has some "over-kill" because only $2(2s + 1) = 6$ should really be needed. Recent applications include a paper by Partha Ghose [43] describing photons and showing the ability to obtain mean Bohmian photon trajectories (2001). This complex analysis essentially resulted in Maxwell equations as before. The prediction was then compared to actual double-slit photon experiments in 2011 yielding sets of similar trajectories [44].
Strictly speaking, Bohmian mechanics derives from the Schrödinger equation for particles with mass, and here we care about the 'idea' of Bohmian trajectories but for massless photons which obey classical electromagnetism. The 2011 experiment used "weak measurements" with only slight changes in polarization (diagonal photons at 45 ° pass through calcite at 42° tilt). The experiment gives an understanding of the distribution of transverse momentum at various focal planes (up to 8.2 meters away from double slits). The resulting sets of trajectories have similar appearance and indicate the slit through which the particles passed because upper and lower trajectories never cross.

For an individual photon transaction, the amplitude and energy of this field is generally too small to collapse into more than one quanta. The space-time development or propagation of A_μ is performed in a way similar to that for classical potentials or fields (iteratively averaging values over neighboring cells as in equation (13)).
There is a long history of treating the vector potential as if it weren't "real." But Maxwell certainly thought it was real. The Aharonov-Bohm effect tends to show it is real in that it can affect quantum mechanical phase differences and shift interference peaks. From the

"non-integrable" form of the equation (14) below, one can write a gauge invariant equation:

$$\delta\varphi \propto \frac{e}{\hbar}\oint_C \vec{A}\cdot d\vec{\ell} = e2\pi\Phi_B/\hbar. \tag{12}$$

where Φ_B is the magnetic flux threading a solenoid. Aharonov and Bohm thought of potentials as fundamental with fields resulting from them by differentiation. Another justification for $\vec{A}$ is that after symmetry breaking for superconductivity, current $\vec{J}$ is directly proportional to $\vec{A}$ (London, 1935); and current is considered to be classically real. If a wavefunction for a superconducting material is written in polar form as: $\psi = \sqrt{\rho}\, e^{i\phi}$, then Josephson tunneling current is $J = J_c \sin(\phi_1 - \phi_2)$ between two superconducting phases. And this phase macro-effect obeys: $\hbar\nabla\phi = q\vec{A}$. Altogether, there is more status for the vector potential now than there was half a century ago. A remaining question is the sub -quantum mechanism for the progression from $\vec{A} \to \vec{E} \to E^2 \to \delta(x)$. That is, there is a progression and emerging to "reality" from an offer wave in sub-quantum "PsiLand" to the classical reality: $\psi \to \psi^*\psi \to \delta(x)$, where the delta function represents final collapse.

6. The Vacuum Machine:

Nature processes physics just fine without us. Unlike a classical vacuum, the quantum Vacuum is a 'machine' that facilitates physical processings. Hopefully, our mathematical physics is at least some sort of isomorphism to what Nature does. We don't know exactly what its mechanisms are, but some sort of 'cellular automata' may offer an elementary perspective. One approach that really inspired me in my practical workplace calculations was the following:

As an example of space-time computation, "How do fields propagate?" and "How can fields be numerically calculated?" A standard numerical technique for solving Laplaces equation, $\nabla^2\phi = 0$, is to divide up the space between two potential boundary surfaces into cells and iteratively update the value of the potential ϕ in each cell over the average of the values in its neighboring cells until numerical results are stable − using some variant of Jacobi or Gauss-Seidel iterations [12]. For example an $n+1$'th iteration from previous n'th values at 3D x,y,z cell i, j, k may look like:

$$\phi*_{i,j,k} = (\phi_{i+1,j,k} + \phi_{i-1,j,k} + \phi_{i,j+1,k} + \phi_{i,j-1,k} + \phi_{i,j,k+1} + \phi_{i,j,k-1})/6. \tag{13}$$

It is highly tempting to say that this is how Nature also does it - **simple sequences of averagings with respect to neighboring space-time 'cells'** − and does so for every different type of field. One might say that each region of space continually calculates and stores and updates and processes these values. Note that cellular atomata (CA) update cell states in terms of states of neighboring cells.[10]

[10]Of course, the idea of a strict lattice of cells is unrealistic, but something approximately like it may still be appropriate.

What are the fields used by Nature? We obviously have the electromagnetic field whose primitive is the underlying EM quantum field or 'photon field' A_μ whose underlying nature is duplicated almost like an ether everywhere throughout space-time. The "Dirac field is a fermion 'spinor field' for the electron quanta having spin-up and down along with the positron-quanta also having spin up and down (the electron-positron field). The EM field and the electron matter field and even the gravitational field can all interact.

An example of this inter-relatedness is a very generalized form of the Schrödinger equation including electromagnetism and weak field gravitation:

$$E\psi = i\hbar\frac{\partial}{\partial t}\psi = \frac{1}{2m}(-i\hbar\nabla - q\vec{A} - mc\vec{h})^2\psi + (q\phi - \mu\,\vec{\sigma}\cdot\vec{B} - \frac{mc^2 h_{oo}}{2})\psi, \quad . \tag{14}$$

where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, $\vec{h} = (h_{01}, h_{02}, h_{03})$ is the metric off-diagonal weak gravitational field (causing the so-called gravito-magnetic or Lense-Thirring Effect), and h_{oo} is the gravito-electric or Newtonian gravity time-metric component (where the usual scalar gravitational potential is $U = -\frac{1}{2}h_{oo}c^2$). We also include the case of a magnetic moment in a magnetic field [11]. So a neutron interferometer could show phase changes due to the earth's gravitational potential and also to a magnetic $\vec{B}$ field applied to one of its pathways. Thus a ψ matter-wave can couple to EM and respond to gravity too. The term $(\hat{p} - q\vec{A}) = m\vec{v} = \vec{\pi}$ is mechanical momentum. $\vec{\pi}$ is physical and gauge invariant.

The classical vector potential, $\vec{A}$, can be considered to have its origin in nearby current flows such as in the Liénard-Wiechert potentials which included 'retarded time' . Then, $\vec{A}$ resembles the 'dragging of electromagnetic space-time' due to moving charges, $\vec{J}$ (and with an appropriate propagation time delay).

$$Li\acute{e}nard - Wiechert:\ \ A_\mu(\vec{r}, t) = \frac{\mu_o}{4\pi}\int \frac{J_\mu(\vec{r'}, t)}{|\vec{r} - \vec{r'}|}\delta(t' + |\vec{r} - \vec{r'}|/c - t)d^3r'dt', \tag{15}$$

For the special case of a time-independent wavefunction without gravity or magnetic moments, an $\vec{A}$ vector potential can cause extra phase shiftings as shown in this solution of the electro-magnetic Schrödinger equation [31]:

$$\psi(x) = \ e^{(iq/\hbar)\int A(x)\cdot d\ell}\,\psi_{(A=0)} = \psi_o e^{i\Delta\varphi}. \tag{16}$$

So, an electron-wave passing around a small solenoid would experience differential phase shifting $\Delta\varphi \sim qA\Delta\ell/\hbar$ along each path $-$ explaining the Aharonov-Bohm electron interference shift. This equation (16) is also a powerful example of how to dovetail or integrate "useful" vector potential fields with "useless" gauge functions. That is, if A has a component which is a gradient of a scalar function, $A_o = \nabla\chi$, then the integral is $\int \nabla\chi \cdot d\ell = \int d\chi = \chi$ due to being a perfect differential. Then the phase change all the

[11] $\mu_e = e\hbar/2m_e$, $\mu_n \sim -1.91\hbar e/2M_p \sim -\frac{2}{3}\mu_p$, and Pauli-matrices σ_i here are unitary 2x2 complex matrices without $\hbar$. Classically, energy $U = -\vec{\mu}\cdot\vec{B}$.

way around the solenoid is zero (i.e., useless). Also, a magnetic field $B = \nabla \times \nabla\chi \equiv 0$ is guaranteed to be null. But, if we instead imagine a winding function, ϕ, about the solenoid, it is multivalued and not a scalar potential and the integral is path dependent. Both of these cases are covered by the same formula.

Since an electron has charge, it is always interacting with the photon (electromagnetic) field [29]. And since a photon effectively possesses an electric field, it can polarize the electron field and also polarize a muon quantum field too. Any of the charged fermionic matter fields feel the electric field of a photon or the charge of any other fermion field. The various fields are often disturbing each other. Also, in the unfortunate term of 'virtual particles,' a photon sometimes becomes a momentary electron and positron and then back to a photon again. And an electron can become an electron and photon and then back to an electron again.

The **strong** field is the color charge gluon field (QCD G_a's based on $SU(3)$) with gluons interacting with the various quark fields. The **weak** field is described by the broken gauge group $SU(2)$ with quanta consisting of W^+, W^- and Z_o bosons. These weak bosons interact with the various lepton fields. The Higgs boson interacts with the weak bosons, the leptons, and the quarks and provides elementary mass to these particles. In a sense, there is a field in the Vacuum for every type of elementary particle (μ, τ, ν_e , ν_μ, u, d, s, c, b, t, ...etc...[sometimes arranged in special 'doublets' and including their anti-particles.]). There is the gravitational field or metric field using the Einstein aether $\equiv g_{\mu\nu}$, various vacuum condensates, fluctuations and dark energy. The superimposed layers of all these fields may be called 'Wilczek's **GRID**' [18]. All of these are special modern examples of Plato's space-time invariant 'Forms' [12]. Nature also holds the values of all the basic physical constants and the rules for all physical operations (duplicated everywhere). It is an impressively very complex machine that suggests an equally impressive deep structure − perhaps with added physical dimensions. This quantum vacuum structure provides sources for all of the possibilities of quantum physics, and the objects of this structure are formed from combinations of the various modes of these fields.

One of the most challenging concepts in quantum mechanics is the apparent existence of a single particle simultaneously taking multiple paths between source and detector. For

[12] Plato (∼*380 BCE*) discussed "Forms" in formulating a possible solution to the problem of universals [19]. His "theory of Forms or theory of Ideas asserts that non-material abstract (but substantial) forms (or ideas), and not the material world of change known to us through sensation, possess the highest and most fundamental kind of reality." " Ontology is the philosophical study of the nature of being, becoming, existence, or reality." Plato argued that "what is real are eternal and unchanging Forms or Ideas." More than a third of modern philosophers advocate a modern form of Platonism. "Starting from the end of the 19th century, a Platonist view of mathematics became common among practicing mathematicians; mathematics is discovered rather than invented. Also note that Heisenberg quantum motivations were in terms of mathematically abstract "Platonic Ideas" over materialism.

example, separate magnetic or electric fields can individually alter the phase of single neutrons simultaneously traversing two spatially separate paths in a carved crystal silicon interferometer. The particle attribute of a full neutron magnetic moment, μ_n, seems to exist in each path at the same time. Also single electrons traveling across a solenoid seem to possess full electron charge, e, in each Aharonov-Bohm pathway. Actuality of particles taking all possible paths at the same time is impossible because their simultaneous summed masses would be infinite, and the summed electron charges would be infinite. So this apparent actualization cannot be classically real. It just behaves as if it were real − hence the inadequate term “simulation.” This concept that “a small part of a wavefunction carries the full mass, spin, charge, energy and momentum” is also discussed in recent papers [14][33].

How can this be? No one seems to have yet volunteered an answer. But it appears clear to me that there are sub-quantum Forms [13] that can be processed by Nature without being classically real. For the Aharonov-Bohm pathways, for example, actual transport of a particle with full electron charge, e, isn’t necessary because the idea of that charge is carried by the “electron” quantum field at all points of space-time [20]. It is readily available everywhere, anytime. There is a growing belief that particles in general really do not exist between source and detector [13][15]; and perhaps particles don’t exist at all but are merely deductions by detectors. There are only quantum fields and waves. In classes on quantum mechanics, we thought we were dealing mainly with particles, but it could have just been quantum waves all along. An electron at a place is really ‘electron-ness’ at that place and time. Prior to ‘collapse,’ an electron can have ‘charge-ness’ and ‘momentum-ness’ over a region of space-time. The words we use are almost all very poor and are being applied in a new realm. Everything needs quotation marks (or italics or underlines) or the phrase “quantum-real” because it is not the same as in our classical world. Most needed words don’t yet exist. But if it is used by Nature, I want to call it a real Form with ‘attribute-ness-es.’

As a recent example of the content of space-time, consider particle-antiparticle colliders producing what might be called “pure energy” which then in turn can output myriad possible output particles of precisely defined types apparently emerging out of the Vacuum itself. Since the earth rotates and orbits, the real 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. Distant light and cosmic rays in astronomy also indicate the invariance of physical laws and constants throughout the universe. There is a beautiful recent plot released by CERN LHC showing quark-mesons produced by the vacuum as seen by increasing total mass of di-muons, $\mu^+\mu^-$ (see Fig. 1 [21]). The spectra of events per GeV begins at left showing spikes 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

[13] Defined as properties and myriad processings that occur to enable the delivery of a single quantum to a detector.

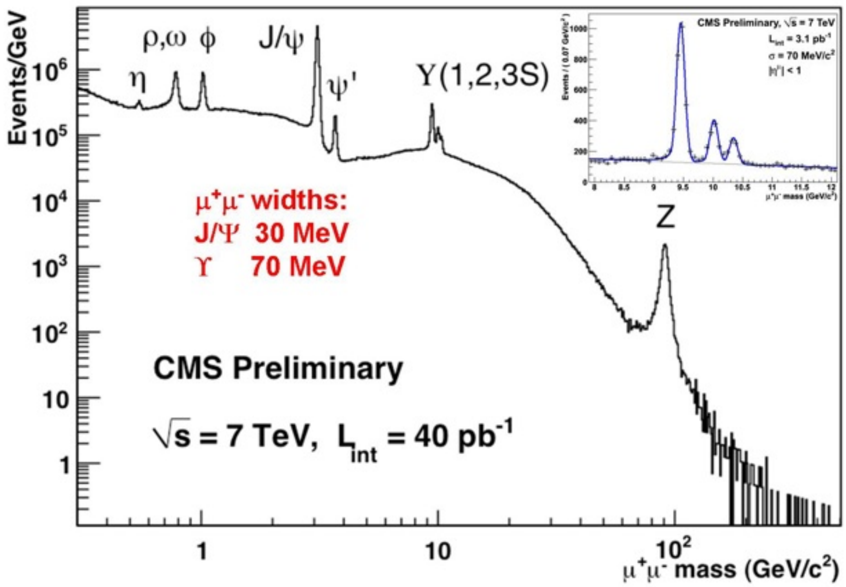


Figure 1. LHC-CMS DiMuon output progression of different resonance states, T.Dorigo, 24 Jan 2011

$s\bar{s}$, 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 boson near 92 GeV. These particles are spewed forth when the Vacuum is stimulated. Their 'Ideas' pre-exist in the fields living in the Vacuum. If a high energy muon- anti-muon collider actually existed, its annihilation would be pure energy that could stimulate the vacuum to produce the particles shown in Figure 1.

All these quarkonia derive from 'elementary' primal quantum fields for 'particles' and their 'anti-particles'. The various quantum fields can be pictured as a stack of planes as in Wilczek's GRID [18]. There is one 2D plane for each quantum field (each fermion, electromagnetism, gravitation $\equiv g_{\mu\nu}$, and much more). But there is a lot of coupling and interacting between these planes which may encourage a more integrative picture. The most popular picture today is that of little spheres or multidimensional "Calabi-Yau" manifolds attached to each point of 4D space-time, $M^4 \times CY$. Above this elementary level in complexity, there is also a realm of composite particles that are not elementary such as the proton, neutron, resonances, multi-quark particles, mesons, nuclei, atoms, and large molecules. These still manage to act as quantum objects.

7. Perspective from Quantum Field Theory:

Quantum Field Theory (QFT) is the mathematical basis of elementary particle physics and can be considered to be encoded by a Lagrangian. A primary difference between QM and QFT is the basic use of creation and annihilation operators to create or destroy "particles." Conserved particles are given in QM, but the number of photons or particle-pairs is variable in QFT. QFT identifies a "wave" with the superposition of an indefinite number of particles.

So, is QFT really based more on particles or on fields? Although there is still a little disagreement (e.g., [28]), a strong majority favor fields as fundamental objects. Nature is made of fields. Quantum fields permeate spacetime, are 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 [15]. 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 [26] 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 wavefunction 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 [26]. 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."

In addition to quantum fields being intrinsically delocalized and unbounded, there is a famous theorem in axiomatic QFT (the Reeh-Schlieder Theorem (1961) [27]) stating that the vacuum is spacelike superentangled! It has has long-distance correlations, and local measurements cannot distinguish between the vacuum and an N-particle (Foch) state. In addition to this, the Unruh effect says that the number of particles depends on the observer (in this case whether the observer is undergoing acceleration). Contrary to naive expectations, it can be said that in a local theory there are no operators counting the particles. In any fixed bounded region in space-time, Separate measurements in space-like separated regions can be correlated in a manner resembling entanglement.

Quantum mechanics offers no explanation for wavefunction “collapse;” it is merely an added postulate due to Heisenberg. If two quanta are involved (bipartite) , then it may involve entanglement. Quantum Field Theory also offers no mechanism for the random, instantaneous, nonlocal, spatial collapse of fields. My guess is that each portion of a single ‘particle’ wavefunction or field excitation is in contact with all other portions as in ‘action at a distance.’ One might say that all portions are ‘entangled’ with each other (although the current definition of entanglement doesn’t venture into sub-quantum mechanics).

Lagrangian View

Lagrangian mechanics (1788) is a reformulation of classical mechanics that turned out to be very appropriate also for quantum mechanics. It is based on an idea of “action” along a path as $S = \int L\, dt$ being stationary or obeying a principle of least action. In specifying a Lagrangian, that means that we are unconcerned about constants of proportionality or additive constants. Finding a solution belongs to the discipline of “calculus of variations” and is facilitated by the basic Euler-Lagrange equations. The simplest introductory case is a choice of $L = T - V$ (kinetic energy minus potential energy). For a constant V, $S \sim \int \frac{p^2}{2m} dt \sim \int p \cdot dq$ as a simpler abbreviated action. In classical mechanics, the action concept is somewhat mysterious, but in quantum mechanical it can be interpreted as “phase” counting by using Planck’s constant, h. In relativistic theory, wave functions are functionals over quantum fields, and the word “particle” can be used for a field that appears in a Lagrangian

A relevant example is a Lagrangian that can yield the Schrödinger equation; and its Lagrangian density, $\mathcal{L}$ can be called a Schrödinger field.

$$\mathcal{L} = -\frac{\hbar^2}{2m}\nabla\psi^\dagger \cdot \nabla\psi - \psi^\dagger U \psi + \left[\frac{i\hbar}{2}(\psi^\dagger \frac{\partial\psi}{\partial t} - \frac{\partial\psi^\dagger}{\partial t}\psi) \to i\hbar\psi^\dagger \frac{\partial\psi}{\partial t}\right] \tag{17}$$

The dagger superscripts refers to “Hermitian adjoint” which results from interchanging rows and columns (for a matrix) along with complex conjugation, *. For many of our cases, $\psi^\dagger = \psi^*$. And later on, QFT introduces the notation $\bar{\psi} = \psi^\dagger\gamma^o$ (for Dirac gamma matrix). The last term in the equation shows the option of a single time derivative or a more symmetrical Hermitian Lagrangian density pair leading to the same final Schrödinger equation after being processed through Euler-Lagrange equations. Since Lagrangian density is not an observable, it may not have to be Hermitian.

One of the first attempts at a quantum field theory for weak beta decay was Enrico Fermi’s 1933 4-point interaction with a Lagrangian term looking like $\frac{G}{\sqrt{2}}(\bar{\psi}_P\gamma_\mu\psi_n)(\bar{\psi}_e\gamma^\nu\psi_\nu)$. These fields would describe a neutron particle decaying into a proton plus an electron (β-ray) and a neutrino in (or anti-neutrino out), $n \to p + e + \bar{\nu_e}$. Unfortunately, the theory failed badly for high-energy with an incorrect prediction of cross sections rising forever with energy. What was needed was the addition of a weak propagator, the discovery of parity violation (1957), and V-A theory [e.g., inserting a $(1-\gamma^5)$ term before the ψ_ν].

The Lagrangian for the whole Standard Model is exceedingly complex [32], and it can take an entire page to write it all out. However, we can simplify it by selecting appropriate terms based on increasing energy and arena. For example, relativistic electromagnetism with classical field source has the following basic Lagrangian:

$$\mathcal{L} = \mathcal{L}_{EM} + \mathcal{L}_{int} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu \Rightarrow \quad \partial_\nu F^{\mu\nu} = J^\mu. \tag{18}$$

And the first term reduces to: $\mathcal{L}_{EM} = \frac{1}{2}(E^2 - B^2)$.
The QED Lagrangian is:

$$\mathcal{L} = \mathcal{L}_{Dirac} + \mathcal{L}_{int} + \mathcal{L}_{EM} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu A_\mu\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{19}$$

The middle term is effectively $-j^\mu A_\mu$, and this plus $\mathcal{L}_{EM}$ gives Maxwell equations as before, $\partial_\mu F^{\mu\nu} = j^\nu$. Dirac theory uses 4-spinors processed by 4×4 complex matrices expressed as combinations of gamma-matrices, γ^μ. Since Dirac spinors can apply to any fermion, the psi's (ψ, $\bar{\psi}$) and mass, m, can have subscripts of $f = e, \mu, \tau, \nu_e, \nu_\mu, \nu_\tau$.

In moving to the electroweak terms at higher energy, the field $F_{\mu\nu} \to B_{\mu\nu}$ where $B_\mu = A_\mu cos\theta_w - Z_\mu sin\theta_w$ and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. That is, we now have to combine the photon A_μ field with the neutral weak boson Z using the Weinberg angle θ_w. But in addition to this, we must also add the W-boson contributions and the chiral lepton doublets (e.g., for the left handed neutrino and electron) and right handed singlets. And then everything becomes complex.

An important observation is that all of the (more than a hundred) terms in the SM Lagrangian represent energies (so that concept is key and common to all the quantum fields). And the physics can be found in terms of least action.

8. Composite Particles

The wave function of the Schrödinger equation has coordinates over all of its configuration space. The hydrogen atom, H, for example, can be expressed as over the positions of both the proton and electron together, $\Psi_H(x_{1e}, x_{2e}, x_{3e}, x_{1p}, x_{2p}, x_{3p})$. In this case its wave function can also be separated and expressed as a product of center-of-mass-system (cms) and relative coordinates, $\psi = \psi_{cm}(\xi, \eta, \zeta)\,\psi_{rel}(x, y, z)$ The wavelength of the CM motion is due to the total mass-energy of the system, $\lambda = h/(M_{total}\,v)$. This is also true for all the composite cases below, although why this is the case is somewhat mysterious.

All atoms and molecules and their nuclei are described by a common wave function in their high-dimensional full configuration space [33]. In some cases (H, He ground state electrons), the particles are also maximally entangled together. There is a rule-of-thumb for when a de Broglie wavelength is due to an aggregate collective mass, and it depends on the effective binding energy (BE) of the system. If the energy of motion of the particle and the interaction energies with the environment are "adequately" below the $|BE|$ of the particle, then the composite system can behave as a single particle with a collective mass. In addition to the momentum-energy waves, this also applies to the spin-statistics theorem. Between 1995 and 1999 it was demonstrated that atoms obey quantum statistics such as ^6Li and ^{40}K for Fermi-Dirac (spin one-half) statistics or ^7Li and ^{14}N or ^{87}Rb for Bose-Einstein statistics.

As examples of this "binding-energy" argument consider the following:
A special case is a Bose-Einstein-Condensate (BEC) of atoms where low temperature and adequately close packing allows neighboring boson atoms to lie inside a de Broglie wavelength of each individual atom (one atom mass). That is, the wave-function is for "the multiply occupied three-dimensional boson field mode [33]." Then the motions of the atoms become highly correlated and move to a collective single wavefunction ground state behaving as a "super-atom."
But, in the case of electrons in a superconductor, very low temperatures allow the formation of Cooper-pairs so that the fermion electrons can pair up like effective bosons. The Cooper-pair binding energy is weak in the meV's, so they retain their "bound" $2m_e$ mass pairing only when very cold.
And consider Compton's observation that some x-rays scatter from atoms through significant angles with no wavelength change. In this case, the photon fails to eject an electron so that the effective Compton mass wavelength is for the whole overall atom instead of for just an electron by itself.

Self-interference is the hallmark of quantum behavior. The first example of bound multi-atom molecules interfering with themselves in a quantum mechanical fashion took place in 1930 when hydrogen (H_2) diffracted from a lithium fluoride crystal surface [23]. But now, Zeilinger (the Vienna group) has done this with large round fullerene soccer-ball shaped molecules composed of 60 or 70 carbon atoms. For the first experiment, C_{60}'s come out of hot ovens, get selected for a speed (say near 117 m/s) and possess a tiny de Broglie wavelength near 2.8 pm (well below the size of a single hydrogen atom!), then get well collimated, pass through a thin silicon nitride grating with spacings of 100 nm and diffract with an angle of only a few arc-seconds. Then the beam or single particles go to a detector 125 cm downstream (ionizing laser detector). The result is shown In Figure 2 where about five peaks of the resulting far-field diffraction pattern can be seen. Functionally, only single balls participate in the diffraction; there are no interactions with other molecules. Note that the diameter of a 60 carbon molecular ball is about 1.1 nm, yet something simultaneously and coherently goes through openings 100 nm apart! Even though their mass is $\sim$ 720 amu's , these are not classical balls. They are quantum objects!

Quantum interference experiments have also been performed with larger molecules such as C_{70} fullerines, larger molecules such as porphyrins and fluorofullerines (2003) and now long molecules like azabenzene (1030 amu). In all cases, clear interference peaks have been seen indicating that single objects interfere with themselves as if they were delocalized

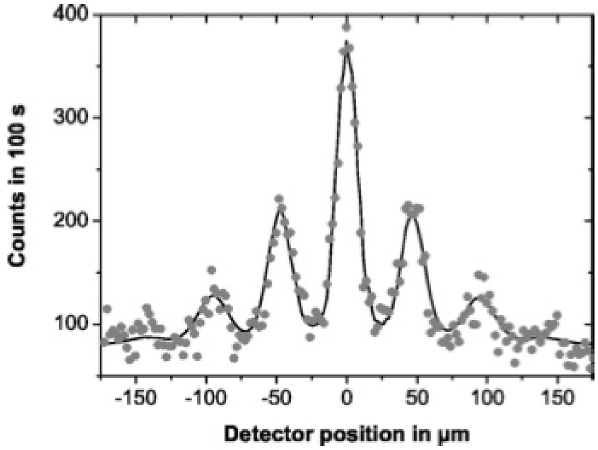


FIGURE 2. Quantum Interference output for Vienna test of C60 Buckyballs. Line curve is theory overlaying experimental dots. Figure from ref. [23]

waves. Experiments can now be performed to show how the visibility of interference drops with imperfect vacuum pressure (e.g., 5×10^{-7} mbar) and increasing temperature of the molecules (e.g., 2000 K). A recent test in 2014 pushed the size limit up to 114 atoms for an atomic weight of 1298 amu and still sees interference effects [37] [14].

In 1960, protons and neutrons were considered to be 'elementary particles.' But we now know them to be amazingly complex objects (not just three quarks). A detected proton spin (for example) must have angular momentum $\pm\hbar/2$, but its sub-distribution within a proton 'bag' is extremely convoluted (part valence quarks, part sea quarks, part gluons, part spin and orbital angular momenta). It is not a simple matter of just adding up three valence quark spins. It is also surprising to note how difficult it is to calculate approximate hadron masses from QCD. What takes us more than 'PetaFlop-Years' of sequential lattice QCD computation, is done easily all the time nearly instantaneously by Nature. Of course, one reason for this is that Nature calculations are fully group parallel, all at once together [15] . Keeping this up over $\sim 10^{80}$ all-identical protons in our universe implies that a proton is also a complex or derived Form of Nature. It is somewhat easy to conjure up because it is one of the few particles that is stable. Its Idea can be said to effectively pre-exist in the vacuum; it is clearly a Form that emerges from fundamental Forms. As entities, hadrons and mesons constantly emerge from the more basic Forms of quarks and gluons. Since much larger many-carbon-atom BuckyBalls (Fullerines) are

[14] That is, tests with precise gratings of $d = 100$ nm on SiNx membrane gratings used high molecular weights for Phthalocyanine, 514 amu, "PcH2" and F24PcH2 ($C_{48}H_{26}F_{24}N_8O_8$) at 1298 amu (114 atoms) with molecular detections to within 10 nm .

[15] It is often said that the enhanced speed of a quantum computer is due to calculating in parallel, but many quantum theorists disagree and point to non-classical alternatives like entanglements, superpositions, interferences, contextualities or even computations in parallel universes [30].

now known to be quantum objects that can interfere with themselves, even these must be complex derived Forms. A question is whether there is an upper limit to this quantumness.

Single neutron 'waves' can travel over multiple paths in a silicon crystal interferometer and interfere with themselves . The two paths can even be at different gravitational potentials (heights) and experience a phase difference between the paths [24]. One path may have electric or magnetic fields that can affect the magnetic moment of the neutron and cause differential phase shifting that can be detected at the output. But again, the neutron magnetic moment is due to very complex internal currents that are hard to calculate, yet Nature does it with ease and can conjure up μ_n anywhere. The fact that hadrons like the neutron have charged quarks (valence udd's) that interact both electromagnetically and through color charge QCD is also intricate and also implies that the layers of the GRID are not independent. Many layers can interact simultaneously.

Antiprotons ($\bar{p}$) were first produced in the lab in 1955 (Nobel in 1959 to Segrè and Chamberlain). They are common in cosmic ray collisions and were created in abundance for the Fermilab Tevatron with $\bar{p}$'s coming from the reaction $p\bar{p}$ collisions from $p + Ir \rightarrow (p + \bar{p}) + p + Ir$. Like the single pairs of 'elementary' particles, the Vacuum knows how to make the next level up composite $p + \bar{p} = (uud) + (\bar{u}\bar{u}\bar{d})$ whenever needed. So, even though there may not be any basic quantum field in this case, there might as well be. Something resembling an emergent effective quantum field for protons (or neutrons) is easily justified.

The next step up is the atomic nucleus. This is a quantum object, and it shows interference effects in scattering experiments. [Examples: angular distributions for center of mass scattering of 17 MeV p on (Zn, Cu, Ni, Co, Fe) with multiple waves for cross sections $d\sigma/d\Omega$ (mb/str); but also projectiles of carbon and oxygen nuclei on carbon and oxygen targets at 24 MeV and later heavy ion elastic scatterings using ^{28}Si, ^{27}Al, ^{20}Ne, ^{58}Ni, $,^{92}Zr...-$; elastic scattering can often be modeled with an optical model ('cloudy crystal ball')]. Note that a typical nuclear density is $\rho_n \sim 0.16\,/fm^3$ (or we might picture a cubic lattice of side $\sim 1.84\,fm$). But the diameter of a proton is 1.76 fm, so there is almost no wiggle room [16]. Yet, somehow, there are nuclear shells and nuclear motions, and hadrons zip around and through the nucleus with ease. These seem to be more like waves that can inter-penetrate each other. They are definitely <u>not</u> spherical stacks of billiard balls!

Returning from quantum field theory back to non-relativistic quantum mechanics, problems using the Schrödinger equation can be broken up into time-dependent calculations and time-independent calculations (TI-Sch). Characterizing the structure of atoms and molecules is largely time independent. The wavefunctions reinforce themselves into semi-stable configurations in space and well defined energies by forming spatially bound standing

[16] RMS electromagnetic radius for pion = 0.66 fm, radius gold = 7.6 fm, radius p = 0.88 fm, but proton charge density is still positive out beyond r = 1.5 fm (although neutron density dies near r = 1.0 fm).

waves. If one asks when and where probabilities form using $\psi^*\psi$, the answer for the time dependent case is generally "at the detector." But covalent chemical bonding occurs because of enhanced electron density between atoms, and that enhancement is due to this "squaring of the wave function" [25].

The simplest case is that of covalent bonding of two hydrogen atoms, H_2 or H_2^+. The most elementary wavefunction is LCAO (linear combination of atomic orbitals) such as two 1S orbitals approaching each other (see equation (4)). The superposition of the two 1S "electron clouds" enhances the electron density between the two nuclei. But this isn't even remotely adequate for bonding. To get bonding, the overlap region has to be enhanced by squaring, (the Born Rule) $\psi^*\psi$ [17]. Bonding occurs without needing any observers probably because the relevant wavefunctions continually and locally reinforce themselves. The situation appears to be consistent with Schrödinger's original claim that the diffuse electron density is given by $dQ/dVol = e\psi^*\psi = \rho(Q)$ (suggesting that the various 'electron clouds' have some reality) [18] That 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.
Note that something similar occurs in the relativistic QED Lagrangian (see eqn.(19)) where a current is identified with a term $e\bar{\psi}\gamma^\mu\psi = j^\mu$ and can be applied to charged fields in general (e.g., for $\psi_{electron}$ or ψ_{proton}). But the interpretation is different in QED, and localization is a problem, and the Lagrangian density is not an observable anyway.

All of these examples are amazing and mind-boggling. What we've been picturing as assemblies of "particles" are actually "wavicles" with waves going around waves with high complexity. A gold atom has 79 electrons in complex "orbits" somehow propagating their motions and keeping track of each other (they can't step on each other's quantum number toes because they are fermions). And then their center of mass also has a wavelength thus allowing quantum interference so that objects of small size can seemingly simultaneously go through double slits separated by large size. It is wasn't experimentally real, it would all be unbelievable. And all of these examples have such difficult mathematical detail as to be nearly uncalculatable. Nature does these difficult things easily and to a degree beyond our abilities.

The quantum information is used and actualized locally without any distant observers being needed. Of course, that statement could be called "metaphysics" beneath Copenhagen quantum mechanics because it is an obvious deduction beneath the realm of direct measurements. But a reminder, again, superposition itself is an obvious deduction; yet

[17] If $\psi^*\psi$ suddenly ceased, you and all your surroundings would suddenly explode.

[18] For example, the potential energy of a 1S orbital for a nucleus of charge Q_N has:

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

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, which is also called one hartree.

no superposition has ever been seen in the classical world. Copenhagen is also logically inconsistent in, for example, claiming that the state function represents the most complete knowledge of a system (that is certainly not a positivist statement). And the statement that quantum mechanics is only relevant to experimental observers makes as much sense as saying that the sun only rises because native peoples rise early to greet it.

9. Sub-quantum Physics

There are no conventional English names for sub-quantum reality largely because of a long tradition that this arena was essentially forbidden by Niels Bohr. Why have ontological names for something that doesnt exist? I've been calling that ethereal world "The Sub-Space-Web [12]," or "The 'Square-Root of Reality," or "Qu-real" or "Pre-Quantum." It might be what Ruth Kastner calls a "Possibilist" [22] world (but each namer has her own unique intentions). It is the physics of the quantum fields of the"GRID" in the Vacuum. Names tend to have a classical chauvinism or bias that isn't quite appropriate. I like the name "PsiLand," motivated by that part of psi that is used by Nature more than by observers.

9.1. **The EM or "Photon" Field.** The photon field is often referred to in terms of the vector potential, A_μ, and that also seems a natural way to refer to the photon wave.
Maxwell's equations should be viewed primarily as an approximate formulation of the physics of the 'photon-field' A_μ as used in PsiLand. These equations were discovered for their classical application in the world of observers, but that is a secondary consequence of their more primitive quantum character. So, we claim "Quantum first, Classical second." The quantum view can be richer than the classical Maxwell view [19]. Maxwell's equations arise as a mean field limit from QED equations where A_μ and the Maxwell field is quantized. And we have already indicated that Maxwell's equations represent single photons (eqn. (10)).

9.2. **The Electron Field.** This field contains values for the electron charge, electron mass, and electron magnetic moment. That also means that an electron traveling a curved path due to an electric or magnetic field responds with its mass-inertia so that the ratio e/m is essentially a coded constant too. When a single electron seems to take multiple pathways at the same time, each path carries a full value of charge and mass. Obviously, that duplication of e and m_e cannot be "classically real."
The Schrödinger equation was written down primarily for electron waves. This is valid

[19] Classical Maxwell equations fail for strong fields and short distances (where vacuum polarization exists). Then quantum electrodynamics has to be used (QED). QED considers the interaction of light and matter and blends special relativity with quantum mechanics. In quantum field theory, A is an operator with different commutation relations than the potential A which is not quantized. Maxwell equations can be derived from QED when we couple a large number of charged particles to their radiation field. Then the two commutators are effectively the same.

for reasonably low kinetic energy where position operators still exist. The electron wave function becomes problematic when electrons are subject to being created or annihilated. Then QED takes over.

The inhomogeneous Maxwell equations (Ampere's law and Gauss' law) relate the electromagnetic field to sources of currents and charges from the realizations of the electron field (mostly primordial electrons and also protons from the infant universe). The creation and annihilation of photons or electron-positron pairs is discussed by the subject of quantum electrodynamics (QED) involving both fields coupled together.
While the A field may be doing the waving for the EM case, what is waving for an electron is a new type of beast: the fermion matter field. The Dirac or Schrödinger equation for an electron in the presence of an EM field involves a term for mechanical momentum $[\vec{p} - q\vec{A}]$. In its operator form as a covariant derivative, so we have $\partial_\mu \rightarrow D_\mu \equiv \partial_\mu + ieA_\mu$. Then, for example in QED scattering problems, we may have a pair annihilation: $e^-(p_1) + e^+(p_2) \rightarrow \gamma(k_1) + \gamma(k_2)$. But electron-positron annihilation could also produce a muon-antimuon pair or even hadron pairs. A photon annihilation product can be considered as pure energy which in turn can output almost anything. Any quantum field possessing a charge attribute will couple to the photon field.

9.3. **A modification to the transactional scenario .** The Transactional Interpretation has a problem with incomplete absorption in an increasingly sparse universe and with "photons that never end" possessing actually energy independently from any consideration of absorption.

We consider the possible hypothesis that there are limits to the previously imagined perfection of quantum mechanics when its fields become very weak. If the amplitude of the wave function has diminished below some tiny critical value, then the quantum machine can no longer transmit or detect the weak quantum information signals. Nearly infinitesimal amplitudes are no longer part of the holistic network of a single quanta. That effectively imposes a distance limit to reliable quantum effects for large travel times [20]. It is supposed that within a realm of adequate strength, the following can happen:
Offer waves can be broadcast from time equals zero out to the limit and can be canceled (effectively instantaneously) by the quantal holistic network in their entirety as needed. This initial wave has field amplitude that can interact with many possible future environments for the purpose of gathering information to make future decisions. Intermediate media can process these signals in their "sub-quantal" ways. Energy is not directly offered or lost by the offer wave ($\psi^*\psi$ is not yet produced). In the absence of any absorbers, cancelation is not needed because the wave self-vanishes at the distance limit. That is, any weak energy beyond the limit is largely not able to be propogated, and some possible residual is in the background noise of space-time.

[20] $\exists\, \epsilon' > 0 \ni |\psi| \leq \epsilon' \Rightarrow |\psi| = 0.$

As in traditional "TI," offer waves encountering absorbers can result in time reversed return information waves that might be acted on.

Depending on return waves or non-returns (again back at $t = 0$), decisions may be made to re-broadcast directed needle rays (plane waves). So waves initially with a spread out factor $f(\theta) \to \delta(\theta - \theta_n)$. Unlike the offer amplitude-waves, these second-pass-waves can result in energy transfer and possess attributes (such as a polarization). These new waves will not attenuate with distance like the offer waves (e.g., $\propto 1/r$). We thus have something like wave and "particle together at the same classical time. We can offer no suggestions for any mechanism for the selection of a final absorber or final direction of propagation.
So, the "photons that never end" are like "photons that have traveled far" in possessing mildly attenuating energy that can gravitate. They should no longer be capable of some quantum properties such as entanglements or contextualities. Unlike spontaneous localizations after some distance, the localization here is determined way back at the emitter with now advanced knowledge of its selected absorber.

9.4. **Photons that have Traveled Far.** Matter or light waves traveling across the universe would eventually have miniscule amplitudes in the midst of overwhelming background noises. It just doesn't make sense that such faint waves could be detectable much less collapse-able. It would be preposterous perfection. In order for "photons to travel cosmic distances, it seems necessary that they no longer be described by spreading Maxwell waves but rather by their more "particle" like aspects while still not be classical particles. This is more like the original ideas of Einstein for "needle rays" (directed Nadelstrahlung, 1917) or de Broglies singularity or soliton type particle core surrounded by some electromagnetic guiding waves. A modern but approximate example of needle rays is laser light (very narrow TEM_{oo} Gaussian beams). But they are electromagnetic, and they do spread and weaken with distance. Solitons of light can exist, but only for special cases in non-linear optical media. So, what is needed is beyond our current conception of quantum mechanics with its limited idea of wave-particle duality. It doesn't suffice to say that QM only describes waves until their final collapse at detection. That may work for some short range within current laboratory experiments but certainly not "long-range." Whatever a particle is, we need it to be able to not just collapse from a wave but for a wave to emerge a particle that can then re-emerge a wave again as it encounters and interacts with matter. An example in biology is that many animals on Earth become dormant to survive dry seasons by "estivating" or forming a protective membrane to keep from desiccating while waiting for the return of wet conditions. The Sub-Space Web may have a feature to maintain long duration identity and energy-momentum conservation in a dearth of interactions.

Let the word **"uniton"** stand for a localized consolidation or re-unification of all the parts or previously active wave regions of a wave-function. It is not yet a formal collapse on a macro-body. So, when a wave has gone too far and gotten too weak, it condenses into a uniton with the intended momentum and energy and quantum numbers of the former wave-function. This entity propagates indefinitely until it interacts with something. It then forms a new quantum wave-function normalized to unity. Because this is a single localized entity, it is incapable of showing quantum effects such as interference.

Is there any evidence encouraging such an entity? We already know from the discussion under the photon wave-function that light from very distant GRBs and neutrinos from supernovae have very little spread in the direction of motion (they stay localized at least in that direction and don't broaden much in the sense of a wave-packet). There are no present astronomical or cosmic tests for complex quantum mechanical effects such as entanglements. The current longest distance quantum test on Earth is the recent optical entanglement-swapping experiment by the Vienna group [45]. Pulses of 404 nm light are used to create two pairs of entangled photons, 0-1 and 2-3. Then a Bell-state measurement (BSM) is performed between 1-2 resulting in the entanglement of photon 0 with photon 3. Photon 3 is then beamed from one Canary Island to a telescope on another 143 km away where its polarization is measured. A verification of the correlation of photon 0 and 3 uses a CHSH Bell inequality S-value test which indeed lies above the value 2 (6-sigma above classicality). Although a very impressive distance, most likely, this test is still below the critical distance for the formation of unitons (and not a cosmic distance).

More thought is needed to elaborate this hypothesis and see if any future verification is possible.

References

[1] Jim Baggott, Farewell to Reality, How modern physics has betrayed the search for scientific truth, Pegasus, 2013.
[2] E. T. Jaynes, "Probability in Quantum Theory" (1996). [PS] prob.in.qm.ps.gz - Probability Theory As Extended Logic bayes.wustl.edu/etj/articles/prob.in.qm.ps.gz?
[3] Steven Weinberg, https://www.facebook.com/pages/Steven-Weinberg/31465905775, or see "Collapse of the State Vector," ArXiv quant-ph 1109.6462 3 Jul 2012.
[4] Matthew F. Pusey, Jonathan Barrett, Terry Rudolph, "On the Reality of the quantum state, Nature Physics 8, 475478 (2012) doi:10.1038/nphys2309, and ArXiv:1111.3328 [quant-ph] 18 Nov 2012, and Nature News 17 Nov. 2011.
[5] Ron Bracewell, The Fourier Transform and its Applications, McGraw-Hill, 1965, e.g., pg. 160.
[6] Ching-Chaun Su, "A Wave Interpretation of the Compton Effect as a Further Demonstration of the Postulates of de Broglie," ArXiv [physics.gen-ph] 2 Jan 2006 0506211 .
[7] Klaas Landsman (Nijmegen) 2009 book sections, "The Born Rule and its interpretation," http://www.math.ru.nl/ landsman/Born.pdf
[8] A. M. Gleason. Measures on the closed subspaces of a hilbert space. J. Math. & Mechanics, 6:885893, 1957.
[9] John G. Cramer, "The Transactional Interpretation of Quantum Mechanics," Reviews of Modern Physics, Vol. 58, No. 3 July 1986, pp. 647-687.
[10] GianCarlo Ghirardi and Raffaele Romano, "Collapse Models and Perceptual Process," ArXiv: 1401.3773 [quant-ph] 15 Jan 2014.
[11] Roger Penrose, The Road to Reality, a complete guide to the laws of the universe, Alfred A. Knopf, N.Y., 2005, pg 512.
[12] David Peterson, "Beneath Quantum Mechanics," pdf, 3/15/2011 − 11/22/12, 42 pgs (unpublished).

[13] Casey Blood, "No Evidence for Particles," arXiv: 0807.3930, also see 0912.2985.
[14] Casey Blood, "Constraints on Interpretation of Quantum Mechanics," http://arxiv.org/pdf/0912.2985.pdf.
[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] Y. Aharonov, S. Popescu, J. Tollaksen, "A time-symmetric formulation of quantum mechanics," Physics Today Vol. 63, no. 11, 27-33, November 2010.
[17] Yakir Aharonov, Lev Vaidman, "The Two-State Vector Formalism: an Updated Review," arXiv:quant-ph/0105101v2 10 Jun 2007. Also see: "Complete description of a quantum system at a given time," J. Phys. A 24, 2315 (1991), and "Time symmetry in the quantum process of measurement," Phys. Rev 134, B1410 (1964).
[18] Frank Wilczek, The Lightness of Being, Mass, Ether, and the Unification of Forces, Basic Books, 2008.
[19] WIK: Platonic Forms: http://en.wikipedia.org/wiki/Theory_of_Forms
[20] Art Hobson, " Electrons as field quanta: A better way to teach quantum physics in introductory general physics courses, 3/4/05, American Journal of Physics, http://physics.uark.edu/hobson/pubs/05.03.AJP.pdf
[21] "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 .
[22] Ruth E. Kastner, The Transactional Interpretation of Quantum Mechanics, The Reality of Possibility, Cambridge University Press, 2012.
[23] Olaf Nairz, Markus Arndt, and Anton Zeilinger, "Quantum interference experiments with large molecules, Am. J. Phys. 71 (4), April 2003 p 319-325. (Free at: http://users.df.uba.ar/mininni/teo2/interferencia_fulereno.pdf
[24] H. Rauch and S. A. Werner, Neutron Interferometry, Lessons in Experimental Quantum Mechanics, Clarendon Press, Oxford (2000). [Includes "COW" experiment (Colella, Overhauser, Werner, 1975)].
[25] Dave Peterson , Explaining S Orbitals and Bonding. pdf, 13 May 2012, 11 pgs, unpublished.
[26] Stephan De Bievre , "Wheres the Quantum?," March 15, 2006 http://www-gat.univ-lille1.fr/∼ debievre/Publications/emch70publicroot.pdf
[27] Helmut Reeh and Siegfried Schlieder, "Bemerkungen zur Unitäräquivalenz von Lorentzinvarianten Felden." Nuovo cimento (10), 22:10511068, 1961. [And see Wikipedia for Reeh-Schlieder Theorem].
[28] David John Baker, "Against Field Interpretations of Quantum Field Theory," July 22, 2008. British Journal for the Philosophy of Science 60 (3):585-609 (2009).
[29] Matt Strassler, "Virtual Particles: What are they?" Of Particular Significance: Conversations About Science with Theoretical Physicist Matt Strassler. October, 2011. http://profmattstrassler.com/articles-and-posts/particle-physics-basics/virtual-particles-what-are-they/
[30] Philip Ball, "Questioning quantum speed," Physics World, Jan 2014 p 38-41.
[31] I.J.R. Aitchison and A.J.G. Hey, Gauge Theories in Particle Physics, A Practical Introduction, Second Edition, Adam Hilger, Bristol, 1989.
[32] Standard Model Lagrangian, W. N. Cottingham, D.A. Greenwood, An Introduction to the Standard Model of Particle Physics, 2nd Ed., 2007, Cambridge, or see summary at: http://einstein-schrodinger.com/Standard_Model.pdf .
[33] H. Dieter Zeh, "The strange (hi)story of particles and waves," March, 2013, translated from German, see: www.eh-hd.de , i.e., http://www.rzuser.uni-heidelberg.de/∼ as3/ , or arxiv:1304.1003 (now v7 Jan 2014, [hist-ph]).
[34] Paolo Tommasini and Eddy Timmermans, "The Hydrogen Atom as an Entangled Electron-Proton System," 17 October 1997, http://arxiv.org/pdf/quant-ph/9709052v2.pdf
[35] I. Bialynicki-Birula, "On the Wave Function of the Photon," Acta Physica Polonica A, Vol. 86 (1994) or http://www.cft.edu.pl/∼birula/publ/APPPwf.pdf or Iwo Bialynicki-Birula, Photon Wave Function, Progress in Optics, http://arxiv.org/pdf/quant-ph/0508202v1.pdf

[36] M. G. Raymer and Brian J.Smith, "The Maxwell wave function of the photon," for SPIE, http://arxiv.org/pdf/quant-ph/0604169.pdf and "Photon wave function, wave-packet quantiation of light, and coherence theory, New J. Phys. 9 (2007) 414, http://arxiv.org/pdf/0708.0831.pdf.
[37] Thomas Juffmann, et. al., "Real-time single-molecule imaging of quantum interference," http://arxiv.org/pdf/1402.1867.pdf Feb 2014.
[38] Cristinel Stoica, "Toward a principle of quantumness," ArXiv:1402.2252 [quant-ph] 10 Feb 2014. Also see: "Can the clicks of the detector provide a complete description of Nature?", 1402.3285 13 Feb 2014.
[39] J. Kovac, et. al., "Detection of Polarization in the cosmic microwave background using DASI," http://arxiv.org/pdf/astro-ph/0209478.pdf, [Degree Angular Scale Interferometer (DASI), located at the Amundsen-Scott South Pole research station].
[40] D. Hansen, et. al. (many), "Detection of B-mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope," 7 Oct 2013 [astro-ph] http://arxiv.org/pdf/1307.5830v2.pdf
[41] R. Hanbury Brown and R. Q. Twiss (1956). "A Test of a New Type of Stellar Interferometer on Sirius". Nature 178 (4541): 1046-1048.
[42] N. Kemmer, "The Particle Aspect of Meson Theory, Proc. R. Soc. A 173 (1939) 91. http://feynman.princeton.edu/∼kirkmcd/examples/EP/kemmer_prsla_173_91_39.pdf
[43] Partha Ghose, et. al., "Bhomian Trajectories for Photons, Phys Lett A 290 (2001) 205-213, or see: http://www.gluonics.com/bohmian/Ghose-2001.pdf
[44] Sacha Kocsis, et. al., "Observing the Average Trajectories of Single Photons in a Two-Slit Interferometer," Science, 332 (6034), 1179-1173, 2011. Blog Discussion: http://scienceblogs.com /principles/2011/06/03/watching-photons-interfere-obs/
[45] Thomas Herbst, et. al., "Teleportation of entanglement over 143 km," (Vienna) ArXiv:1403.0009 [quant-ph] 28 Feb 2014, refers to previous Ma, X.-S, et al. "Quantum teleportation over 143 kilometers using active feed-forward." Nature 489, 269-273 (2012). [also see 1205.3909 and 1205.2024].

10. Appendix:

ENTANGLEMENT is an increasingly important topic describing the non-classical linking of two or more objects such that the quantum state of a constituent requires that of the others even when spatially separated. A definition is that an entangled state cannot be written as a tensor product of states, so "the probability of nding two particles with given positions or momenta is not simply the product of the single particle probabilities."
Beginning students of quantum mechanics solve the hydrogen atom problem by describing its wave function as a product of center-of-mass-system (cms) and relative coordinates. The relative distance between the proton and electron is $\vec{r} = [x, y, z] = \vec{r_1} - \vec{r_2}$, and the Coulomb potential is $\propto e^2/|\vec{r_1} - \vec{r_2}|$. The electron and the proton **are entangled**, and a proof of this can be found in [34] [21].
Similarly, the two electrons in the ground state of the helium atom are entangled. Simply placing two electrons in the same place (or by having their wavefunctions overlap) produces their entanglement because of the need to anti-symmetrize. The two electrons of the helium ground state also have spin-entanglement. The entangled singlet state is $|\uparrow\downarrow - \downarrow\uparrow\rangle$.

HANBURY BROWN TWISS (HBT) EFFECT:

In 1956, Astronomer Robert Hanbury Brown and mathematician Richard Q. Twiss (HBT) [41] measured the size of the star Sirius using light across the disk of the star as

[21] An exact analytical solution for entanglement is quantified using a density matrix formalism which describes correlations in the electron-proton motion. It is not conventionally transparent.

seen by signal intensities on two large parabolic detectors spaced some 'd meters apart. The photo-detections have to be fast acting, and both signals have to be processed at the same times. The star diameter $D \sim R\lambda/d$ can be approximated easily assuming classical coherent light across the star (which is of course unreal) using just a geometrical path difference for constructive phase difference within about half a wavelength of the light. For Sirius (α-Canus Majoris A) the diameter is $D = 3.4D_{\odot}$ at a distance of R = 8.6 light-years seen by photomultiplier detectors spaced about $d \sim 6$ meters apart (with capabilities to hundreds of meters). HBT two-detector intensity interferometry at that time could only detect blue light from hot stars and ultimately processed 32 stars prior to abandoning this Australian built technology in 1974. The actual physics uses chaotic light of different frequencies requiring quantum photon optics theory to obtain boson bunching effects (e.g., Roy Glauber 1962). It came as a surprise that photons of different frequencies could interfere this way. Light beams from the edge of the star travel to the two detectors using uncrossed and also crossing rays, and bosons cannot distinguished between these choices. The distance from the detectors to the ray-crossing point is about 8000 km or trillions of wavelengths ($h = \lambda(R/D)^2 = d^2/\lambda$). After 1974, astronomers switched to conventional amplitude interferometry.

Questions: Electromagnetic Photons: Since a photon path can traverse an assembly of optical glass items, it is obvious that a photon wave must be electromagnetic in the Maxwell sense. But to what degree are the quantum Maxwell field and the classical Maxwell field the same. Should the quantum EM wave be considered ghost-like in a separate "reality" prior to its approaching classicality by interactions or large numbers (as in a laser beam?). Is there a quantum simulation world? In addition, if a photon is A, to what extent are the derivatives taken to give E and B also. And are these realized truly as energy E^2 and B^2 for traveling waves?
Because of "photons that never end" one cannot rely on the existence of absorbers or measurers for a complete physical description of quantum mechanics. It must be that contexts and entanglements are attributes that are optional depending on whether interactions with absorbers have occurred or not. In the absence of absorbers, photons still represent real energy that in its totality can still gravitate. But, somehow, angular momentum still has to be conserved (and that must happen at emission ??).

Note that the special relativity case of de Broglie's relation as an operator is $\hat{p}_\mu = i\hbar\partial_\mu$ – which covers both the momentum operator and the Hamiltonian energy operator together in one SR formula [22]. That is, for timelike metric signature $\eta_{\mu\nu} = (+---)$ [23] [as in most texts on QFT, and with $c = 1$], and $\partial_\mu = (\nabla, \partial/\partial t)$, then:

(20) $p^\mu \equiv (\vec{p}, E)$, $p_\mu = \eta_{\mu\nu}p^\nu = (-\vec{p}, E) \to i\hbar(+\nabla, \partial/\partial t) \Rightarrow \vec{p} \to -i\hbar\nabla$, $and\, E \to i\hbar\partial/\partial t$.

[22] If the spacelike signature, $\eta_{\mu\nu} = (-+++)$ is used instead, then we get $\hat{p}^\mu = i\hbar\partial^\mu$.

[23] Unlike "Our" Euclidean space, SR uses a reference of "Light." So, any photon has a metric distance $ds^2 = c^2dt^2 - dr^2 = 0$ and 'particle' distances with $v < c$ are 'off the light cone" (or "null cone").

GAUGE THEORY FOR ELECTROMAGNETISM

DAVE PETERSON

ABSTRACT. Usual discussions of gauge theory begin with gauge symmetry transformation invariance and then work towards the nature of physical interactions which maintain symmetry. The gauge principle of local gauge invariance yields the **forms** of interactions. When carried over to relativistic Dirac Lagrangians, it allows for the form of full Maxwell equations with Lorentz force! But even in the simplest electromagnetic (EM) case, it is not readily apparent why invariance using gauge functions should imply the <u>actual</u> existence of a useful non-zero EM-vector potential, A.

It may be more appropriate to begin with the solution to the EM Schrödinger equation, $\psi(x) = \psi_o(A=0)\, e^{(iq/\hbar)\int A(x)\cdot d\ell} = \psi_o\, e^{i\varphi}$ and then consider a gauge transformation modification $A' = A + \nabla\chi$. That yields $\psi(x) = \psi_o e^{i\varphi}\, e^{iq\chi/\hbar}$ because $d\chi = \nabla\chi \cdot d\ell$ is a perfect differential. Then establish consistency with physics by showing that a scalar field χ will not alter classical fields E or B or the sense of the Schrödinger equation. The general form for the electromagnetic quantum phase alteration $(q/\hbar)\int A(x)\cdot d\ell$ over some path ℓ covers **both** the pure gradient gauge function term (χ) with path-independent effect and real non-integrable non-gradient phases which do depend on path together under one concept. The 'useful' vector potential can be formed from a **directional derivative** of the right kind of phase function, $\hat{n}(\hat{n}\cdot\nabla)\varphi$ rather than a gradient. The important 'Aharonov-Bohm' (AB) effect can experimentally detect a net path-dependent φ phase shift and is insensitive to the hypothetical and useless χ scalar field. The AB effect applies to particles with mass and electric charge that can couple to the A field − even in those cases in which no net magnetic field, B, is present (for example, the exterior field of a solenoid).

1. INTRODUCTION

The unfortunate and misleading use of the term 'gauge' in modern quantum field theories goes back to its original 1918 introduction by Hermann Weyl. He attempted to unify electromagnetism with general relativity gravitation by introducing local gauge transformations of 'length' or scale and used a mathematical formulation still applicable today. The word 'gauge' was commonly used as the setting of a standard of length, linear distance or dimension (such as the separation between railroad tracks)[1]. In field theories such as electromagnetism, a 'gauge standard' or gauge selection or 'gauge fixing' restricts the functional freedom of field potentials, e.g., the sensible Coulomb gauge $\nabla \cdot A = 0$ will

Date: December 16, 2011.
email: davepeterson137@gmail.com. Paper updated to December 10, 2012, April 9, 2013.

[1]60% of the world's railways use a standard gauge of 1.435 meters. A wheeled cart gauge near this width goes back to the Bronze age.

not alter the values of B or E fields. This, of course, has nothing to do with lengths but merely represents a standard 'constraint' over a sum of values. This gauge choice enables a useful conventional form for equations over potentials without introducing any observable physical consequences. The main purpose of a gauge choice is to simplify calculations.

Weyl's 'gauge scale' concept quickly died, but his mathematics became appropriate to quantum mechanical phase. So, in quantum mechanics, the word gauge might be replaced with 'phase' instead. When we speak of gauge transformational invariance under $A' \to A + \nabla\chi$, the gauge freedom of the χ 'gauge function' can apply to the classical field or the quantum phase functional. For phase, $\theta = e\chi/\hbar$, a 'modulo' (2π) radians is implicit. This enables a compact Lie circle-group U(1) understanding for the gauge function (or phase function). With the modern use of the gauge principle, electrodynamics is now called a gauge field theory (perpetuating the confusion between gauge versus phase). Since the benefit of its use in deriving the Standard Model of particle physics, there is strong interest in gauge freedom and gauge invariance − the 'relevance of irrelevance' [1].

2. Classical Electromagnetism Background

A classical charged particle in modest motion in an electromagnetic field experiences a 'Lorentz force' given by

$$F = q(-\nabla\phi - \partial A/\partial t + v \times [B = \nabla \times A\,]\,), \tag{1}$$

where $\vec{A}$ is called the vector potential, and q or e is the charge on a particle − a conserved quantum number. After Heaviside's 1885 reworking of Maxwell's equations [5], $\vec{A}$ was considered to be just a mathematical convenience lacking any physical reality. But it is now known to have some important and basic degree of physical reality because it can change the phase of an electron through the Aharonov-Bohm effect [2]. The Lorentz force can be expressed using a velocity dependent potential [3] given by $U = q\phi - qA \cdot v$. Then the simplest Lagrangian form of kinetic energy minus potential energy , $L = p^2/2m - U$ gives a 'canonical momentum':

$$p = \partial L/\partial \dot{q} = \partial L/\partial v = \mathbf{mv} + \mathbf{qA} \tag{2}$$

Note that in 1856 Maxwell used to refer to A as the 'electromagnetic momentum' [2] [11]. In the relativistic case, one can begin with another velocity dependent form called the Minkowski Force $K_\mu = qF_{\mu\nu}u^\nu$ giving canonical momentum $p^\mu = mu^\mu + qA^\mu$. $F_{\mu\nu}$ is the 4x4 relativistic field strength tensor.

Classical electromagnetism is based on Maxwell's equations which include Ampere's Law, Faraday's Law of induction, Gauss' Law, and 'no-poles' − or in SI units:

---

[2]− a term I still like for the product qA.

$$\nabla \times B = \mu_o J + \epsilon_o \mu_o \frac{\partial E}{\partial t},\ \nabla \times E = -\frac{\partial B}{\partial t},\ \nabla \cdot D = \rho,\ \&\ \nabla \cdot B = 0 \tag{3}$$

Expressing the fields E and B in terms of potentials comes largely from the vector identities that the divergence of a curl is zero and that the curl of a gradient is zero. So $\nabla \times B = 0$ implies that B could be a curl of something, $\mathbf{B} = \nabla \times \mathbf{A}$. And Faraday's law

$$\nabla \times E = -\frac{\nabla \times A}{\partial t} = \nabla \times \frac{\partial A}{\partial t},\ or\ \nabla \times \left[E + \frac{\partial A}{\partial t}\right] = 0 = \nabla \times (\nabla \phi). \tag{4}$$

So, $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial\mathbf{t}$. Note that these potentials come from the "homogeneous Maxwell equations" (the ones without sources or currents). The minus sign back-reaction against the magnetic field in this equation and in Faraday's law is called Lenz's Law. Now, in the $B = \nabla \times A$ equation, A itself could also consist of some 'core A' due to local currents and some extraneous gradient field $\nabla\chi$. The $\nabla\chi(x)$ field would be incapable of contributing to a classical field and would be redundant or irrelevant and express what is called 'gauge freedom.' In the special case of a field free region, $B = 0$, A might have a form resembling $= \nabla\chi$, and the 'gauge function' $\chi = \int \nabla\chi \cdot d\ell$. If one chooses to add a $\nabla\chi(x)$ term to A, then one is immediately also required to perform a compensation change to the voltage potential ϕ as well. That is, if $A' = A + \nabla\chi(x)$, then preserving the electric field, E, too requires:

$$E = E' = -\nabla\phi - \frac{\partial A}{\partial t} = -\nabla\phi' - \frac{\partial(A + \nabla\chi)}{\partial t} = -\nabla\phi' - \frac{\partial A}{\partial t} - \nabla(\frac{\partial\chi}{\partial t}) \tag{5}$$

which means that the scalar potential $\phi' = \phi - \partial\chi/\partial t$. So, the $-\partial\chi/\partial t$ compensates for the $+\nabla\chi$ term; a voltage potential compensates a vector potential. This can be expressed by relativistic 4-vectors as $A'_\mu \to A_\mu + \partial_\mu\chi$ to preserve the anti-symmetric $F^{\mu\nu}$.

Perhaps the most common classical gauge in electromagnetism is the Lorenz gauge condition, $\partial_\mu A^\mu = 0 = \nabla \cdot A - \epsilon_o\mu_o\partial\phi/\partial t$, [3] which still allows free fields χ such that $\Box^2\chi = 0$.

In the above, we said that we can freely modify the potentials of electromagnetism and still describe the same EM field under the transformations:
$A' = A + \nabla\chi(x)$, and $\phi' = \phi - \partial\chi/\partial t$. In his book on Quantum Mechanics [24], Cohen-Tannoudji points out that $q\nabla\chi(\vec{r}, t)$ can be considered a non-physical transformation of momentum, $\vec{p}$. That is, in the Hamiltonian formalism, the conjugate momentum of position, $\vec{p} = m\vec{v} + q\vec{A}$ is different from the mechanical momentum $\vec{\pi} = m\vec{v} = \vec{p} - q\vec{A}$. It is $\vec{\pi}$ that is a physical quantity which preserves its value under two different gauges and $\vec{p}$ which is non-physical and depends on the arbitrarily chosen gauge. That is:

[3] From the Danish physicist Ludvig Lorenz not Hendrik Lorentz but actually dating back to 1888 by George FitzGerald. However, for whatever historical error, it is a frequent tradition in physics to name things after their second or third discoverer rather than the first. So, perhaps the spelling doesn't matter. Maxwell himself preferred the Coulomb gauge $\nabla \cdot A = 0$,

$\vec{\pi} = \vec{\pi'} = \vec{p} - q\vec{A} = \vec{p'} - q\vec{A'} \Rightarrow \quad \vec{p'} = \vec{p} + q\nabla\chi$.
"In the Hamiltonian formalism, the value at each instant of the dynamical variables describing a given motion depends on the gauge chosen." This observation carries over to quantum mechanics as well. But, in QM, χ more clearly resembles a 'phase.'

A common expression for the electromagnetic Lagrangian density would show its kinetic term $p^2/2m$ added to its field, where $\mathcal{L}_k = (p_\mu - qA_\mu)^2/2m$ and a free gauge field part:

$$\mathcal{L}_{EM} = -\frac{F_{\mu\nu}F^{\mu\nu}}{4} = \frac{(E^2 - B^2)}{2} \tag{6}$$

This $\mathcal{L}_{EM}$ portion is invariant under the transformation $A'_\mu = A_\mu + \partial_\mu\chi$. But some added mass-like term such as $-A_\mu A^\mu/2$ would not allow invariance [6]. The lack of this term is consistent with a massless photon.

It is noted that most physics texts present logical derivations of the mathematics encompassing physical concepts, but they generally refrain from interpretation of the mathematics and often even from providing definitions of terms.[4] The foundations of physics tend to be just givens. This is true not only for quantum mechanics and modern physics but also for some of classical physics as well. This is due to a well warranted paradigm of avoiding speculation not necessitated by direct experiment. Next year's experiments might overthrow previous paradigms causing embarrassment. Electromagnetism is well over a century old, so we should be increasingly able to answer basic questions such as "What is a magnetic field?" We began the discussion above with the existence of the Lorentz force $F = qv \times B$ guiding the form of the canonical momentum. This equation is also used as a practical **definition** of B – but it puzzles (or at least should puzzle) college students who notice that a B field in one direction and a particle velocity in another direction results in a force in a third perpendicular direction. In oversimplified terms, a magnetic field is the shear of the swirling of the vector potential field, A. But then we must ask the further question, "What really is A?"

In general relativity, there is a similar phenomenon called 'gravitomagnetism' or the now experimentally verified 'Lense-Thirring effect' where a spinning mass tends to drag inertial frames around with it (the GRT vector potential $A \simeq ch_{io}$ – an off-diagonal time-space term of the linearized metric tensor and with units of velocity). A distant particle approaching a spinning black hole will be deflected by this inertial rotational frame-dragging similar to that of a Coriolis force, $F = 2m\omega \times v$. [5] Similarly, a rotating electrical current flow drags its vector potential, A, around with it. So, the Lorentz force could be viewed as the motion of a charged body in a space with a shearing of the magnetic vector potential

[4]Nature is the owner of definitions in physics, and we gradually reveal them through theory and experiments. Basic definitions are provided in this paper at the end.

[5]Also seen in deflected winds coming from the equator and sometimes causing hurricanes or for curving ocean currents

field − a progressively changing 'frame of reference' [6]. We view its trajectory against our inertial background, but an electron has a charge portion that also sees the rotating A-field as well [7]. However, for E&M, it seems to be incorrect to refer to a dragging 'velocity'. The dragging of currents seems to affect charged particle momentum, and momentum is more basic in quantum mechanics than velocity.

The fundamentally basic 'Aharonov-Bohm' (AB) effect from 1959 [2] says, "contrary to the conclusions of classical mechanics, there exist effects of potentials on charged particles, even in the region where all the fields (and therefore the forces on the particles) vanish." One of these effects is that an electron phase shift can be produced due to the presence of a vector potential $\vec{A}$ without any accompanying classical magnetic field, $\vec{B}$. This has been demonstrated by observed shifting of interference fringes in a variety of experiments (e.g., an electron beam splitting and then rejoining around a long thin solenoid). The external A field is dragged in space by current flowing around the solenoid and with the trajectory of one e-beam and against the motion of the other beam.

Consider a short distance segment Δx for an electron path along a vector potential. Perhaps a "proper quantum momentum frame" of reference is one which is dragged along with the vector potential by an amount $p_{em} = eA$ [7] also called the "field momentum" (Thomson, 1904) or electrodynamical momentum. The phase change over Δx for $A = 0$ is $\Delta\varphi \simeq \omega\Delta t = 2\pi\nu\Delta x/v_\phi$. Let mechanical or 'kinetic' momentum in an inertial frame with no magnetic vector potential field present be given by $p_m = mv$. There is a canonical (generalized, conjugate) momentum which is most relevant to the important deBroglie wavelength of quantum mechanics. That has the form:

$$p_{con} = mv + qA = \frac{h}{\lambda} = \frac{\partial L}{\partial v} = -i\hbar\nabla = \hbar k. \tag{7}$$

[in contrast, the Hamiltonian stresses the kinetic momentum and writes it as $p_m = mv = p - qA$]. We now consider the difference between these two momenta with the A field off versus A field on:

$$\Delta\varphi \simeq 2\pi\nu\Delta x(\frac{1}{v_\phi} - \frac{1}{v_{\phi'}}) = 2\pi\Delta x(\frac{1}{\lambda} - \frac{1}{\lambda'}) = \frac{\Delta x}{\hbar}(p_m - p_{con}) \tag{8}$$

$$\Delta\varphi \simeq \frac{eA\Delta x}{\hbar} \sim \frac{e}{\hbar}\int \vec{A}\cdot d\vec{x} \tag{9}$$

The use of the vector potential A in electromagnetism has been horribly undertaught and underappreciated. As one example, we think of a plane electromagnetic wave as a crossed E and B field propagating over distance and time. If an old high vertical radio broadcasting antenna has an up and down quickly oscillating electrical current J, this current flow can be considered as dragging 'electromagnetic space' up and down with the current. When

[6] but you can't transform it away.

calculated using the Lorenz gauge, this A-field propagates outwards at the speed of light because $\Box^2 A = -\mu J$. If 'up-and-down' is in the unit vector $\hat{z}$ direction and we view the field in the unit vector $\hat{x}$ direction, we see:

$$\vec{A} = \frac{\mu_o}{4\pi}\int \frac{\vec{J}d^3x}{r} \rightarrow \quad \tilde{\mathbf{A}} = \hat{\mathbf{z}}\mathbf{A_o}\sin(\mathbf{kx} - \omega\mathbf{t}), \tag{10}$$

where $\omega/k = c$. Then, derivative fields B and E are calculated from the vector potential A field:

$$B = \nabla \times A = -\hat{j}\partial A_z/\partial x = -\hat{j}\ A_o \cos(kx - \omega t),\ and, \tag{11}$$

$$E = -\partial A/\partial t = -\hat{z}A_o(-\omega)\cos(kx - \omega t) = +\hat{z}A_o\cos(kx - \omega t) \tag{12}$$

Notice that

$$\hat{E} \times \hat{B} = \hat{i} = \hat{x},\ and\ |B| = k|E|/\omega = \frac{2\pi\, 1\, |E|}{\lambda\, 2\pi\nu} = \frac{|E|}{c}. \tag{13}$$

This vertically polarized 'up-and-down' propagating A field provides a much easier (and perhaps more fundamentally physical) picture than the crossed E and B fields.

Historically, potentials weren't observables. And another reason for its low usage is that there are few easily calculated closed form analytic solutions to EM problems in terms of A. Using E and B enables easier calculations. But in digital computing over meshes or finite elements, it is sometimes easier to calculate V and A first and then derive E and B from the potentials second. Perhaps Nature also does it that way.

3. Quantum Mechanics and Electromagnetism

Now move from the classical world to the quantum world. For non-relativistic quantum mechanics, the Schrödinger equation expresses energy conservation, $E = p^2/2m + V$ using operators for energy and momentum, $\hat{E} = i\hbar\partial/\partial t$, $\hat{p} = -i\hbar\nabla$. If this was extended to include EM interactions with Hamiltonian type $p \rightarrow mv = p_{canonical} - eA$, one would expect to see something like:

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = i\hbar\frac{\partial\psi}{\partial t} = E\psi \rightarrow \quad \frac{1}{2m}(-i\hbar\nabla - qA)^2\psi = (E - q\phi)\psi \tag{14}$$

This is an 'electromagnetic Schrödinger's Equation.' For the case $\phi = 0$, the solution to this equation is [4]

$$\psi(x) = \ e^{(iq/\hbar)\int A(x)\cdot d\ell}\ \psi(A = 0) \tag{15}$$

This form covers vector potentials which are "real" and also for unreal additions like the gauge term $\nabla\chi$. In general, the 'real A' use is 'non-integrable' so that the phase in the exponent **depends on which path is followed**. For the case of a closed curve path, this just gives the enclosed flux $\Phi = B\mathcal{A}$ by Stokes' Theorem. $\Delta\varphi = q\int A(x)\cdot d\ell/\hbar$ is a functional which converts multi-dimensional or vector input into a scalar output, the phase (most modern texts choose natural units where $c \equiv 1\, and\, \hbar \equiv 1$). It is very important to note that the phase difference $\Delta\varphi$ is only defined for a choice of path. That means that it is <u>not defined as a scalar field</u> in the sense of the chi scalar field. The vector potential can be expressed as a decomposition $A' = A + \nabla\chi$. If we consider just the irrelevant pure gradient $\nabla\chi$ portion by itself, then the integration becomes trivial (integrable) because [4]

$$\nabla\chi \cdot d\ell = \frac{\partial\chi}{\partial x}dx + \frac{\partial\chi}{\partial y}dy + \frac{\partial\chi}{\partial z}dz = d\chi, \tag{16}$$

with a solution depending only on the end points: $\psi(x) = \psi(\chi = 0)\exp(iq[\chi(x) - \chi(-\infty]/\hbar)$. So, turning on some $\chi \neq 0$ field would introduce a phase shift on ψ. If $\chi = \chi(x)$ varies with x, then the phase shift varies with local location. The gauge concept is that any local phase changes that occur with equation (14) requires that the following three mutually compensating things have to occur together:[7] [8]

$$\psi' = \psi\, e^{-iq\chi(x,t)/\hbar},\;\; A' = A + \nabla\chi(x,t),\;\; \&\;\; \phi' = \phi - \frac{\partial\chi(x,t)}{\partial t}. \tag{17}$$

Many books and articles discuss gauge invariance of electromagnetism and say that having a local phase change leads to the existence of a vector potential field, A. These discussions are rarely clear or convincing. What they show is not an A field deduced from changed wavefunction phase, ψ, as much as just total self consistency of simultaneously having $\psi \to \psi' = \psi e^{ie\chi/\hbar}$ with $A' = A + \nabla\chi$ and potential $\phi' = \phi - \partial\chi/\partial t$ (which won't affect E and B fields), and the use a special 'covariant derivative $D = \nabla - ieA/\hbar$ in the Schrodinger equation− all together. Change the momentum operator from ∇ to D, change other terms to primes, expand, and the chi goes away (e.g., see [16]). This self-consistency implies that local phase invariance of the intent of the Schrodinger equation goes along with the use of A and ϕ in the proper EM Schrodinger equation.

The demonstration is somewhat subtle. We can first consider a single differentiation and don't even need to take $\nabla^2 \to D^2$ [16].

$$D\psi' = e^{iq\chi/\hbar}[\nabla + iq\nabla\chi/\hbar - (iq/\hbar)(A + \nabla\chi)]\psi = e^{iq\chi/\hbar}(D\psi). \tag{18}$$

Similarly, for $D_t = (\partial/\partial t + iq\phi)\psi$, we have:

$$D_t\psi' = e^{iq\chi/\hbar}[\partial/\partial t + (iq/\hbar)\partial\chi/\partial t + (iq/\hbar)(\phi - \partial\chi/\partial t]\psi = e^{iq\chi/\hbar}D_t\psi \tag{19}$$

The classical Maxwell invariance of A' and ϕ' are also gauge invariances of quantum mechanics. And this carries over to the relativistic Dirac QED equations with space-time

[7] Vladimir Fock, Fritz London, 1926-1928

[8] The gauge transformation using the label chi is fairly conventional. But other labelings include: $A \to A + gradient\, of\, (\Psi, \Lambda, \chi,\, or\, f)$. Quantum phase adjustment is $\alpha = \theta = q\chi$ using the same chi as for A. The frequent metric convention has $g_{oo} = \eta_{oo} = +1$.

4-vector A_μ as well. The space and time parts are coupled together.

Some popular books mention that changing a voltage reference is an example of gauge invariance. But getting an additive constant from $d\chi(t)/dt$ means that χ ramps up uniformly forever $-$ what kind of real function does that? (the winding function). Changing potentials by a constant amount is an example of a global transformation. Similarly, a constant gradient field for $\nabla\chi$ also cannot exist (as a standard vector gradient of a scalar field). These gauge shifts accompany the A and ϕ fields, but we mainly care about phase shifting due to the A and ϕ fields, not from the χ fields. We care about going from ∇ to D and not much from D to D'. With respect to the usual Schrodinger wave, the A(x,t) potential field changes its phase locally!

So, backtracking a little, to even before the previous discussion, let ψ satisfy a Schrodinger equation without electromagnetism. Now modify ψ with $\psi = \psi e^{i\theta}$ and also let $del = \nabla$ become $\nabla - ieA/\hbar =$ operator D. Then D on ψ gives the older $\nabla\psi$ plus new terms:

$$D\psi' = [\nabla\psi + i\psi\nabla\theta - ieA\psi/\hbar]e^{i\theta} \tag{20}$$

The two new terms go away if $i\nabla\theta = ieA/\hbar$, or $A = \hbar\nabla\theta/e$.

But, a special new trick is needed that A is a vector in some special direction (the direction of source current flow, $\hat{J}$), and we only consider the rate of change of theta in that direction. So theta is not considered as a scalar function that provides a gradient $-$ and that is good because the curl of a gradient is zero (no possible B field makes A pretty useless). Then we consider only a "directional derivative" $\vec{A} = (\hbar/e)\hat{x}\nabla_x\theta = (\hbar/e)d\theta/dx$ or change of wavefunction theta phase over a length delta x is: $\Delta\theta = eA\Delta x/\hbar$ like in the Aharonov Bohm effect. This coordination between A and θ is in agreement with the solution to the EM Schrodinger equation (15): $\psi(x) = e^{(iq/\hbar)\int A(x)\cdot d\ell}\,\psi(A=0)-$ which says that the dot product in the integrand locks in directions of path along the unit vector of the vector potential. This is different from the gauge function concept of the gradient of a pure scalar function, χ because of equation (16). We've gone from a useless and highly unphysical gradient of a phase function to a directional derivative $\hat{A}\nabla_{\hat{A}}$ (the unit vector direction of A or current). There should be a nicer math way of saying this and it should be better known. Note that many of the Gauge Theory discussions avoid speaking of the **sources** of the relevant potentials $-$ the vector and scalar potentials are simply assumed and are placed into useful formulas like the electromagnetic Schrödinger equation. For electromagnetism, the sources of potentials are currents and charges. Fields are derivable from these sources using the 'Liénard-Wichert' formulas from electromagnetism. Including source and U(1) phases into relevant pictures for electromagnetism can be done [e.g., 'Little Circles Model' [17]].

A quantum field theory (QFT) is conveniently expressed in a Lagrangian formulation. Unlike the Hamiltonian, the Lagrangian is a Lorentz scalar [8]. Also, important conservation laws easily follow from the symmetries of the Lagrangian density, through the use of Noether's theorem − for example finding conserved currents, $\mathcal{J}_\mu$. The Lagrangian density for a quantum mechanical particle can be given is several forms. One is the non-relativistic Schrödinger equation:

$$\mathcal{L} = i\hbar\psi^*\dot{\psi} - \frac{\hbar^2}{2m}\nabla\psi^* \cdot \nabla\psi - V\psi^*\psi + \mathcal{L}_A \tag{21}$$

We derived the last two compensations in equation (5). The first phase change equation provides a meaning for χ − it is a scalar field which couples to a change in QM phase. But it is usually more relevant to use the Dirac QED equation for a fermion

$$\mathcal{L} = \bar{\psi}(iD_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{22}$$

where one could substitute the canonical derivative $\nabla \to \nabla + ie\vec{A}$ or $D_\mu = \partial_\mu + ieA_\mu$ so that $iD_\mu = i\partial_\mu - eA_\mu$. The 'slash' notation could be used to show Dirac gamma matrices bases, but that is understood. These are the Lagrangian forms one considers for the following concept: The 'gauge principle' advocates that general fundamental interaction terms for weak and strong forces as well as electromagnetic interactions can be stated by performing 'local gauge transformations on the kinetic energy terms in the free Lagrangian for all particles.'

The Dirac equation (27) above also applies to a 'free' fermion quark but with more complicated gauge fields ${A^i}_\mu$. The shape of the covariant derivatives for non-Abelian $SU(N)$ cases over fields resembles:

$$D_\mu\phi_k = (\partial_\mu - ig{A^i}_\mu\lambda^i)\phi_k \tag{23}$$

where g is the appropriate coupling constant. For SU(3), there are 8 phase angles, and the λ matrices (or Gell-Mann matrices) are 3x3 matrices. The wavefunction would be a quark colour triplet wavefunction which transforms by phase changes $\exp(i\alpha(x)\cdot\lambda/2)$. The division by two refers to the fermion spin-one-half property that it takes 720^o or two full rotations to return to the identity transformation. Fundamental gauge field interactions are with matter fermions.

For the Weak interactions using the symmetries of the Lie group SU(2), write:

$$\mathcal{D}_\mu = \mathcal{I}\partial_\mu + ig\mathbf{T}\cdot\mathbf{W}_\mu(\mathbf{x})/2 \tag{24}$$

where $\mathbf{T}$ is the tau's (τ_1, τ_2, τ_3) or Pauli matrices for three phase angles θ, W includes three gauge fields (or Yang-Mills fields), $\mathcal{I}$ is the 2x2 unit matrix identity, and these operate on an isotopic spin doublet 2-vector or 2-spinor ψ. The $\mathbf{T}\cdot\mathbf{W}$ is a 2x2 complex matrix. To also include a $\times U(1)_Y$ effect as well would entail adding another term to the covariant derivative, $ig'YB_\mu/2$ with a new coupling constant g'. This is almost as if in addition to the idea of EM momentum, there might also be weak-isospin and hypercharge momenta as well (and perhaps there is also a 'colour momenta'). These act 'as if' they could 'drag' the matter momentum phase along with the new field flows.

We could let a fermion-doublet wavefunction phase transformation look like $\psi' = \exp(i\theta \cdot T/2)\psi$, but for Lie Groups we prefer to work with the much simpler Lie Algebra infinitesimal angle transformations instead, so $\psi' = \exp(i\eta \cdot T/2)\psi$ where $\eta = (\eta_1, \eta_2, \eta_3)$ small angles. Then $W' = W + \delta W$ or $W'^\mu = W^\mu - \partial^\mu \eta(x) - g[\eta \times W^\mu]$. So, in the non-Abelian case, we no longer just get the 'throwaway' gradient addition but also now have a term $[\eta \times W]$ added on as well. Unlike the Abelian case (e.g., $U(1)$), we have a coupling of the small angle change with the gauge potential.

The Higgs particle at $m_H = 126\,GeV$ was announced by the CERN LHC on July 4, 2012. It is the quantum of the Higgs field which gives mass to the gauge bosons and fundamental fermions. The Higgs mechanism enables effective mass while preserving gauge symmetry thus violating the conditions for the zero mass Goldstone theorem to apply [18]. In other words, the relevant fields (call "A") interact with the complex field ϕ to give an interaction term $\propto \phi^*\phi A^2$. Symmetry breaking displaces potential to $\phi = H + \mu$ so that $\phi^*\phi A^2 \to \mu^2 A^2 + ...$ resembling a mass term like $m^2A^2/2$ in the quantum field theory.

EW theory with group $H = SU(2)_L \times U(1)_Y$ undergoes reduction via the Higgs field to a $U(1)_{EM}$ subgroup (diagonally embedded in H, [Witten]). Note that a 2×2 write-out of a Lagrangian term: $(ig\mathbf{T} \cdot \mathbf{W}_\mu(\mathbf{x}) + ig'YB_\mu)$ contains diagonal terms $(gW^{(3)} \pm g'B)$ proportional to the photon A and neutral weak boson Z^o [10].

There are complex wavefunction fields for the three cases of interest: U(1), SU(2), and SU(3). A complex field means that complex phases are involved which could be labeled by their gauge functions. We know χ for U(1). There are two of these phases for SU(2) which might be labeled χ_a or in some texts as α_a [10] forming a column vector $\vec{\chi}$ or $\vec{\alpha}$. For QCD or SU(3) with its 8 color gauge fields, there are three phases. Then, just as our EM examples above, when forming the covariant derivative, there will be an added term like $\partial_\mu \vec{\alpha}$.

4. Intuitive concepts without math:

We are all familiar with scalar fields such as the collection of all temperature values all over the globe. We are also familiar with the concept of gradient such as a 15% grade downhill on a mountain road. The gradient of the scalar temperature field is a map of all the arrows pointing from high temperatures to low temperatures (the directions of maximum change along with lengths or labels of their strengths $-$ or in other words, a vector field). A magnetic field is the shearing of a vector potential field to give local curling or swirling. For swirling fluids, this can be seen by placing small logs or pinwheels or balls in the fluid and seeing if they turn or rotate. If they do, then curl the fingers of your right hand in the direction of rotation with the thumb pointing out either up or down accordingly. This conventional "right-hand-rule" thumb direction is said to be the direction of the curl and magnetic field. The "vector-potential, A" fluid field of a loop of electric current flow is dragged along with the current. The more loops, the stronger the A-field. The curl of this field is the magnetic B-field.

An interesting fact about gradient fields of scalar functions is that it is impossible to give them a swirl-curl. For example, there is no way to label temperatures so that the heat flow will swirl around a circle. As an example of the challenge, picture say eleven bricks spread out uniformly along the rays of a circle from some central hub. Let the brick at zero degrees east (or positive x-axis) be at 100 degrees (Celsius) or just enough to boil water. The brick just above that is at 90 degrees, then 80 degrees and so on around the circle to the last brick near freezing. So, there is a circular temperature gradient from the East up to the North and around the circle. But the last brick at zero degrees freezing cold coincides with the hottest brick at 100 degrees! That can't happen! The shearing-swirl or curl of a gradient of any scalar field is zero! Or the magnetic field of the gradient of a gauge function (chi) doesn't exist (has a zero value). If the only thing we care about in the world is swirling heat flows, then the scalar temperature field has no utility.

Now, if a gradient in temperature does exist, something else will automatically happen. Heat will flow from hot to cold over time. A space gradient causes a time flow. If the gradient were due to scalar pressure fields instead, then there will be a resulting fluid flow of wind or water. This is an attempt to make the temperature or pressure differences go away towards equilibrium. The space differences cause a compensating flow of something over time.

There is a somewhat exception to the usual rule that $B = \nabla \times A = 0 \Rightarrow A = \nabla\chi$ in the case of the A-field outside a solenoid. In this interesting and highly relevant case, A does look like the gradient of χ, and the phase or $\chi \propto \phi$, the angle around the solenoid. The phase and chi are lifted in going around like an helical-ramp (annular-helix, helical staircase, 'split-lock-washer', or spiral ramp 'Riemann surface' for log z $\sim i\theta$ [15]). It does look like an *almost exception* scalar field except that it isn't a function because it is multi-valued and also that it cannot be defined at $\rho = 0$. The integral of $A \cdot d\ell$ is zero around any closed path except a path all around the solenoid itself. The problem, of course, is that any path around a hole is not simply connected [19] and cannot have a true-gradient field by an extended Green's theorem.

The present discussion is about electricity and magnetism and how they affect the waves associated with traveling particles. There is a funny effect called Lenz's Law which attempts to bring changes in field strengths back into equilibrium. If a magnet approaches a closed coil of wire or any conductor, the increasing magnetic field strength near the wire will induce a current flow in an opposite direction so as to create a new magnetic field from the coil which reduces the over-all felt strength there − a back field. An increasing magnetic field goes along with an increasing vector flow field, A. The rising A field causes an electric field to be produced in the wires in the opposite direction − the infamous minus sign of Lenz's law (the minus sign in Faraday's equation (3)). For the ghostly gauge field (chi) a positive gradient for A results in a minus change in the ghost gauge function on the electric field E. Now the vector potential A and the scalar voltage V or ϕ are real because they can affect phase in the AB effect. These ghostly gauge function fields may or may not have real existence– we will never directly measure them. They go along for the ride and do similar funny things like the potential fields do.

So far, it is just a basic fact of Nature that 'particles' have an accompanying wave-like field of some sort (the 'deBroglie Wave' or 'matter wave' or 'quantum information field'). The packing of wave peaks together in space (wavelengths) codes for what we call momentum, and the packing of wave peaks together in time (frequency) codes for what we call 'mass-energy.' Maybe mass-energy really is just frequency. Perhaps space-time is like a 'computer' which can read and process these codes. We have been talking about the vector potential field $\vec{A}(\vec{r})$ as if were some fluid flow. That is an old picture that was used by Maxwell in the 1860's along with his pictures of gears and wheels in space − the 'molecular vortex model.' This picturing was discarded by Oliver Heaviside in 1884 along with the elimination of the A's and V's from Maxwell's equations. Just leave behind the safe mathematics without taking the risk of interpretation. Relativity (how moving observers see moving things) made the picture more difficult, but the mathematics stayed the same. It turned out that Maxwell's equations already contained relativity in disguise (Lorentz Invariance). Einstein's first relativity paper was on the electrodynamics of moving bodies (it was rejected as a PhD thesis). But, if we are careful about reference frames, it is not too wrong to consider the A field as fluid motion as seen by eyes that can see the flow (bodies with charges). In that case, an electron beam that gets split around a solenoid and then interferes with itself can be explained by each beam about the coil traveling either with or against the fluid motion so that it waves either get expanded or compressed. Then the phase difference causes a shift in the center interference fringe at a detection screen. Now, a century after Heaviside, we are re-inserting the potentials back into Maxwell's equations with new appreciation. But there are still many student who know very little about the magnetic vector potential, A.

Are the gauge functions, chi(x,t), really ghostly? Or do they represent background fields that we have decided not to care about? If our labs on earth really were traveling through an A_o constant fluid flow field as background, we would never know about it. We would only take space and time changes on the field (curl and time derivatives) and count those as real. A_o wouldn't contribute to B or E or even to the AB effect (because it alters both electron paths the same causing no net phase shift). Should non-contributing background fields be called chi-fields? It is hard to imagine any chi field that could have a constant gradient or a constant time change. Or perhaps, there is a background chi field which has wavelengths and oscillates on the 3-D mesh of space. And, as mentioned above, it is possible to have a chi field such that $\Box^2\chi = 0$ – implying that chi can propagate at the speed of light. What would be the consequence of taking such an idea seriously?

5. Does Gauge Theory Lead to Electromagnetism?:

There are common statements that the gauge principle of local gauge invariance implies Maxwell's equations and the existence of the photon, A_μ. One meaning is that if gauge theory can derive a no-sources solution $\partial_\nu F^{\mu\nu} = 0$, then that means $\Box^2 A_\mu = 0$ which is a wave equation for A which can have a photon quanta. The claim that gauge theory leads to Maxwell's equations of electromagnetism sometimes implies a restricted form of the equations such as those for free space outside of media (no polarizability, no magnetizability) and often with no charge or current sources. But the application to the Dirac QED Lagrangian can give the form of the full Maxwell equations including the Lorentz force. It is important to clarify these claims.

Once more, the general Maxwell equations are:

(G) Gauss' Law: (or Coulomb's Law) which in S.I. units is $\nabla \cdot \vec{D} = \nabla \cdot \epsilon_o \vec{E} = \rho$,
(N) No-poles: $\nabla \cdot \vec{B} = 0$, [but $\nabla \cdot (\nabla \times \vec{A}) = 0$, so $\vec{B} = \nabla \times \vec{A}$] ,
(F) Faraday's Law: $\nabla \times \vec{E} = -\partial B/\partial t$, or $\nabla \times \vec{E} + \partial B/\partial t = 0$, and
(A) Ampere's Law: $\nabla \times \vec{H} = \vec{J} + \partial \vec{D}/\partial t$.

(L): There is also an interaction with charges called the Lorentz force $\vec{f} = \dot{p} = q(\vec{E} + \vec{v} \times \vec{B})$[9]. This is equivalent to having a "velocity dependent potential" $U = q\phi - q\vec{A} \cdot \vec{v}$ where qv could be treated as a current j so that an $\vec{A} \cdot \vec{j}$ could go into a Lagrangian as an interaction term.

Now $\vec{D} = \epsilon_o \vec{E} + \vec{P}$ and $\vec{B} = \mu_o(\vec{H} + \vec{M})$; but gauge theory mainly cares about EM for free space outside of solid media. So the relevant Ampere's Law would be stated as:
(A':) $\nabla \times \vec{B} = \mu_o \vec{J} + \mu_o \epsilon_o \partial \vec{E}/\partial t$, or $\nabla \times \vec{B} - c^{-2} \partial \vec{E}/\partial t = \mu_o \vec{J}$.

To get a wave equation, we note that in (F) we could have included $0 = \nabla \times \nabla\phi$ and we need to impose the Lorenz gauge condition $\nabla \cdot A + \mu\epsilon \partial\phi/\partial t = \partial^\mu A_\mu = 0$, and then we get for the vector potential or 4-vector potential:

$$\nabla^2 A - \mu\epsilon \frac{\partial^2 A}{\partial t^2} = -\mu J, \quad and \quad \Box^2 A_\mu = -\mu J_\mu. \tag{25}$$

[9] The original Maxwell equations included all of these along with Ohms Law $J = \sigma E$ and the continuity equation $\nabla \cdot \vec{J} = -\partial\rho/\partial t$.

The other view of a wave as crossed E and B fields simply comes from this A-wave using $\nabla \times A = B$ and $E = -\partial A/\partial t$.

Quantum Field Theory (QFT) prefers to work with relativistic notation using the anti-symmetric field tensor: $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ which might be called a '4-curl' of A. The homogeneous Maxwell equations (F and N) follow directly from this definition, the cyclic sum $\partial_{[\mu}F_{\nu\lambda]} = 0$.
One can show that the relativistic electromagnetic Lagrangian leads to the inhomogeneous Maxwell's equations (Gauss, Ampere) if the A potential is varied while leaving the current density constant [20]. That is:

$$\mathcal{L} = \mathcal{L}_{EM} + \mathcal{L}_{int} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu \Rightarrow \quad \partial_\nu F^{\mu\nu} = J^\mu. \tag{26}$$

Just having the free field first term without the external current J would only give $\partial_\nu F^{\mu\nu} = 0$ with no sources. Now, this can also be expressed as [4] $\partial_\nu F^{\mu\nu} = \Box^2 A_\mu - \partial^\nu(\partial_\mu A^\mu) = 0$. But, the Lorenz gauge condition then just leaves the wave equation for A; and that in turn has plane wave solutions. If quantization is applied, then we have the photon.

In deriving physical theory from the gauge principle, one usually first begins heuristically with the QM wave function $\psi(x,t)$ for the Schrödinger equation and then transform it with local phase invariance (arbitrary phase modification at each point of space-time, $\chi(x,t)$: $\psi(x,t) \rightarrow \psi'(x,t) = \psi(x,t)e^{i\chi(x,t)}$. One then immediately finds that the simple Schrödinger form is not preserved because the altering of the phase has to be due to the influence of some kind of force field. That is, the ∇ of the ∇^2 term for the momentum operator has to be modified to a form $\vec{D} = \nabla - iq\vec{A}$ as mentioned in a previous section. And then along with $\psi \rightarrow \psi'$ we also need $D \rightarrow D'$, $\vec{A} \rightarrow \vec{A}' = \vec{A} + \nabla\chi$, $V \rightarrow V' = V - \partial\chi/\partial t$. Then the shape of the EM-SCH equation is preserved as shown in equation (14), and the vector field A^μ that guarantee's local phase invariance is called the 'gauge field.' This derivation does not insure that q is the proper EM charge and that A^μ is the familiar EM vector potential. Showing that requires quantum field theory. And one must show that the A gauge field is not the trivial (e.g., $B = 0$) or integrable A like the gradient of a scalar field [4]. We want useful phase from proper A to be path dependent like that seen in the Aharonov-Bohm effect (1959).

The Dirac equation for a charged particle in an EM field is [21]:

$$H\psi = [-ic\hat{\alpha} \cdot (-i\hbar\nabla - e\vec{A}) + \hat{\beta}mc^2 + e\phi]\psi = i\hbar\partial\psi/\partial t \tag{27}$$

Where $\hat{\alpha}, \hat{\beta}$ are 4×4 matrices and ψ is a 4-spinor. In terms of gamma matrices where $\gamma^o = \beta$, $\gamma^i = \beta\alpha^i$, (and $\hbar = c = 1$), it is more common now to write $(i\gamma^\mu D_\mu - m)\psi = 0$. An electron four-vector current density can be written as $j^\mu = e\bar{\psi}\gamma^\mu\psi$ with a conserved Noether current $\partial_\mu j^\mu = 0$. This is a really key concept, so let's elaborate a little here:

A continuity equation for fluids with density ρ is: $\partial\rho/\partial t + \nabla \cdot \rho v = 0$. But our equation above for j^μ seems to be missing a velocity term, v. The Eherenfest correspondence principle for quantum mechanics versus classical mechanics is:

$$\langle v_x \rangle = d\langle x \rangle/dt = (i/\hbar)\langle \hat{H}x - x\hat{H} \rangle \sim \langle p_x \rangle/m \tag{28}$$

For the case with the Dirac equation Hamiltonian, $\hat{H}$ above, eqn. (27), we would obtain $\langle v \rangle = c\hat{\alpha}\psi$. Multiplying the free field Dirac equation from the left by $\psi^\dagger$ indicates that the terms in the continuity equation are a probability density of $\rho = e\psi^\dagger\psi$ and a current of $j_i = ec\psi^\dagger\hat{\alpha}_i\psi$. Now $\bar{\psi} = \psi^\dagger\gamma^o$ and $\gamma^o = \beta$ where $\beta^2 = \gamma^{o2} = 1$. That means that we can insert a one into the $\jmath_i$ equation as $1 = \gamma^o(\gamma^0)^{-1} = \gamma^o\beta$ so that:

$$j_i = ec\psi^\dagger(1 = \gamma^o\beta)\ \hat{\alpha}_i\psi = ec(\psi^\dagger\gamma^o)(\beta\alpha_i)\psi = ec\bar{\psi}\gamma^i\psi. \tag{29}$$

There is a velocity term which is just the speed of light which is conventionally set to unity (c = 1).

A more common expression for the electromagnetic Lagrangian density would show its kinetic term $p^2/2m$ added to its field, where $\mathcal{L}_k = (p_\mu - qA_\mu)^2/2m$ and a free gauge field part:

$$\mathcal{L}_{EM} = -\frac{F_{\mu\nu}F^{\mu\nu}}{4} = \frac{(E^2 - B^2)}{2} \tag{30}$$

This $\mathcal{L}_{EM}$ kinetic field portion is invariant under the transformation $A'_\mu = A_\mu + \partial_\mu\chi$. The overall QED Lagrangian can then be written as: $\mathcal{L} = \mathcal{L}_{Dirac} + \mathcal{L}_{int} + \mathcal{L}_{EM} =$

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu A_\mu\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{31}$$

The middle term is effectively $-j^\mu A_\mu$, and this plus $\mathcal{L}_{EM}$ gives Maxwell equations as before, $\partial_\mu F^{\mu\nu} = j^\nu$.

For gauge invariance of Dirac QED when $\psi(x) \to e^{i\theta(x)}\psi(x)$, forcing $eA^\mu \to eA^\mu + \partial^\mu\theta(x)$ introduces the covariant derivative $D_\mu = \partial_\mu + ieA_\mu$. This adjustment of the derivative is called the minimal coupling prescription. The crucial point is that substituting $\partial_\mu \to D_\mu$ into the Dirac Lagrangian automatically gives the interaction jA term $\mathcal{L}_{int}$ which agrees with the term for the "vertex" of quantum electrodynamics. And then one can say that the q and A really are the familiar charge and vector potential.
That is:

$$\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \to \bar{\psi}(i\gamma^\mu(\partial_\mu + ieA^\mu) - m)\psi = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - e\bar{\psi}\gamma^\mu A^\mu\psi. \tag{32}$$

And that last term is the JA interaction term corresponding to QED.

Combining this with $\mathcal{L}_{EM}$ gives the inhomogeneous Maxwell equations!. Imposing the gauge principle on the free fermion Lagrangian leads to the interacting field theory of QED [10]! and "all of electrodynamics [22]."

Is this argument now clear? As before, we are introducing the modifications $A_\mu \to A_\mu + \partial_\mu\chi$ and $D_\mu = \partial_\mu + ieA/\hbar$ which eliminates an unwanted $\partial_\mu\chi$ term from the transformed Lagrangian. But also again, why does A have to be non-zero? The gauge principle doesn't force the existence of a non-zero A field. We simply have previous knowledge of electrodynamics and Feynman QED and a claim that some field must have altered the local phase. Yang and Mills abstracted from an already understood EM, so their gauge prescription is a given. A more modest claim [23] is that the principle of local gauge invariance determines the **form** of all interactions. It 'reflects a deep relationship between the universality of the various interactions, conservation of the vector currents, and the existence of the interactions themselves.' Another statement is that "there is ultimately no compelling logic for the vital leap to a local phase invariance from a global one [4]. But, it is still considered basic because leads to the form of interactions.

6. The 'Useful' A Field:

The scalar field, χ discussed so far is called a gauge function. It can look like a phase if multiplied by $q/\hbar$, so that χ can be expressed as $b \times phase$ where $b = \hbar/q$. What we wish to propose now is a possibly 'useful phase' φ as in a wavefunction multiplication by $\exp(i\varphi)$. That is, define phase $\varphi = (q/\hbar)\int A_{core}(x) \cdot d\ell$ for the useful (non-gradient) A field. Even though this is not a scalar field in the sense of chi(x), it could be well defined as a phase coupled with a vector direction $\hat{n} = \hat{\Sigma J}$ (like a flag in the wind of the current sources). It is again understood that this phase can be altered by a global addition of any arbitrary angle; only its derivative will be of value. The purpose of this new useful phase is to lead to a useful vector potential which can give a magnetic field or an AB effect (unlike any phase from χ). Most of the time, the gauge transformation addition $\Delta\vec{A} = \nabla\chi$ is 'useless' and won't correspond to physically meaningful cases.

We now wish to consider the 'useful' vector potentials as having the form $\vec{A} = b\hat{n}\nabla_n\varphi$, a **'directional derivative'** vector instead of a gradient of a scalar function of varying phase. A directional derivative is a gradient projected onto a certain direction by direction cosines, $\hat{n}\nabla_n\varphi \equiv \hat{n}(\hat{n}\cdot\nabla)\varphi$. The direction $\hat{n}$ unit vector is supplied by the net background or source currents $\Sigma\vec{J}$. There are also cases in which the gradient and the directional derivative appear to agree. There may be special phase functions and unit vector basis directions which are aligned with the direction of choice so that the gradient and the direction are parallel.

Exterior Solenoid Example: One of the most important instances of the latter case is the vector potential field surrounding a long solenoid of radius R and current I (and let the axis be in the z direction). This A field will not cause any net B field, but it is still real because it can cause an Aharonov-Bohm phase interference effect when electron beams split around the thin solenoid (and an interference build-up from an ensemble of single individual electrons). Then, for this exterior or 'outer' solution we have:

$$\vec{A}_{out} = \vec{A}_\phi = K\hat{\phi}/\rho, \quad where\, K = \mu_o n I R^2/2 = B_o R^2/2. \tag{33}$$

The gradient of a phase field in cylindrical coordinates is:

$$\nabla\varphi = \hat{\rho}\,\partial\varphi/\partial\rho + (\hat{\phi}/\rho)\,\partial\varphi/\partial\phi + \hat{z}\,\partial\varphi/\partial z.$$

We wish to find the phase field that will yield the A vector above. Since this is a case of 'useful' A, we have the directional derivative:

$(\hat{\phi}\cdot\nabla)\varphi = \nabla_\phi\varphi = (1/\rho)\partial\varphi/\partial\phi = K/b\rho$, so phase is $\varphi = \varphi(\phi) = K\phi/b$.

The phase field builds up as one moves in the phi direction around the solenoid helix and keeps on doing so past a full revolution. In this particular case, it is also true that $\hat{\phi}\nabla_\phi\varphi = \hat{\phi}(\hat{\phi}\cdot\nabla)\varphi = \nabla\varphi$; the gradient agrees with the directional derivative. BUT, this case is not a true gradient because the phase wraps around the solenoid like a multivalued helix which is not a single valued 'function' on a simply-connected region. So, we calculate as if the gradient worked, but it is really an illusion.

Since $d\vec{r} = \hat{\rho}d\rho + \hat{\phi}\,\rho d\phi + \hat{z}dz$, we also have a phase integral
$\varphi = (1/b)\int \vec{A}\cdot d\vec{r} = (1/b)\int(A_\phi\hat{\phi})\cdot(\rho d\phi) = \int K d\phi/b$.
Then $b\,\nabla\int\vec{A}\cdot d\vec{r}/b = K\nabla\phi = K\hat{\phi}/\rho = \vec{A}_\phi$. The integral represents a phase, and the gradient of the phase is the vector potential (as was mentioned in equation (15)).

Interior Field of Solenoid: Another interesting example is the interior solution of the solenoid where a net non-zero uniform magnetic field $\vec{B} = \nabla\times\vec{A}$ The vector potential A field for this case is given by:

$$\vec{A}_{in} = \mu_o In\rho\hat{\phi}/2 = B_o\rho\hat{\phi}/2 = b\hat{\phi}\nabla_\phi\varphi = \hat{\phi}(b/\rho)\,(d\varphi/d\phi) = \hat{\phi}\frac{\hbar}{q\rho}\frac{d\varphi}{d\phi}. \tag{34}$$

Then the phase is given by $d\varphi = B_o\rho^2 d\phi/2b$ so that $\varphi = \varphi(\rho,\phi) = B_o\rho^2\phi/2b$.
If we were to take the gradient of this phase field, we would obtain not just a phi component but also an unwanted radial component as well. That is, $b\nabla\varphi = \vec{A}_\rho + \vec{A}_\phi$.

$$\nabla\varphi(\rho,\phi) = \hat{\rho}B_o\phi\rho/b + \hat{\phi}B_o\rho^2/2b\rho = (B_o\rho/b)[\hat{\rho}\phi + \hat{\phi}/2]. \quad so\ \vec{A}_{in} = b\nabla_\phi\varphi \neq b\nabla\varphi. \tag{35}$$

The usual gauge assumption of taking the gradient of phase fields does not apply. And the curl of the gradient of φ is $B = B_z = (\nabla\times\vec{A})_z = 0$! rather than the actual $B = B_o$. A gradient assumption clearly does not work here. We have to use directional derivatives of a phase field instead.

We can again also utilize A phase by calculating :
(36)

$$\vec{A}_{in} = \hat{\phi}\nabla_\phi\left[\int\vec{A}\cdot d\vec{r} = \int(\vec{A}_\phi\hat{\phi})\cdot(\hat{\phi}\rho d\phi) = \int\frac{B_o\rho^2 d\phi}{2} = \frac{B_o\rho^2\phi}{2}\right] = b\frac{\hat{\phi}}{\rho}\frac{\partial\varphi}{\partial\phi} = \frac{B_o\rho\hat{\phi}}{2}.$$

[where $b\nabla_\phi(\int\vec{A}\cdot d\vec{r}/b)$ cancels the b's].

So, does that extra unwanted radial vector get in the way of showing that the idea of directional derivative is consistent with the electromagnetic Schrödinger equation? Well, analysis from just looking at the form of the D term won't work any more because of the extraneous radial term. But, the Laplacian is the divergence of the gradient, $\nabla^2\psi = \nabla\cdot\nabla\psi$. So the new $D^2\psi' = D\cdot D\psi'$. And the dot products impact the radial terms: $\hat{\rho}\cdot\hat{\phi} = \hat{\phi}\cdot\hat{\rho} = 0$

and $b\nabla\varphi \cdot A_\phi = bA_\phi{}^2$. The net result is that equation (15) still works.

It may be argued that Gauge Theory should have been presented in a more physically meaningful way:
a). Local Phase Field: Rather than introducing an arbitrary worthless non-physical locally varying electromagnetic phase field as a ‘function (meaning single valued on simply connected region), allow it to be physically plausible and multi-valued (like the ramping or winding function).
b). Electromagnetic U(1) gauge functions should not be formed from ugly and useless gradients. The space-time substrate instead must use directional derivatives oriented according to net background currents ($\hat{\phi}\hat{n}$). So, we have $\hat{\phi}\nabla_n\varphi$.
c) Begin with the usual SCH equation essentially having A = 0. Transform using the compensating or cheat term ‘$0' \to 0 + \hat{\phi}\nabla_n\varphi = 0 + \vec{A}$. The useful A and the useful phase go together (the sources cause A which causes the varying phase), $p \to p + eA$.
d) The form preserving transformation with directional derivatives also eliminates the net appearance phase terms. In usual gauge theory, an ugly useless gradient of phase term is contrived to eliminate its appearance. That fails to provide any useful potential A.
e) The QM phase for “matter waves is different from the U(1) phase for EM. For a plane wave, the gradient of psi does yield $\nabla(\vec{k}\cdot\vec{x}) = \vec{k}$. Matter waves represent scalar amplitude information about particles (including massless photons) and include other fields in terms of their interactions with matter (e.g., qA). EM waves are for vectors (or 4-vectors).

In Summary: Why does the gauge principle work? I think the answer is found by working backwards. Classical EM and the electromagnetic Schrödinger’s equation both use EM momentum term eA. The solution to the EM SCH equation is equation (15). This equation works for both the phase derived from a directional derivative and also for the chi using a gradient − it applies redundantly for both concepts. If we pretend that we are ignorant and don’t know what a local phase transformation should look like, then equation (15) will apply to both the real directional derivative and also to the gradient concept. So, under general ignorance of the phase, we can justify both and get $A' = A_{real} + \nabla\chi_{unreal}$. The unreal chi just goes along for the ride.

One current point of confusion is what to do with the wave equation for A where the Lorenz Condition is applied effectively also leading to a wave equation in the chi phase term $\Box^2\chi = 0$. Could this be an example where gradients are meaningful?

Loosely speaking,
1) Possible A fields: There are real $\vec{A}$ (like the Liénard-Wiechert potential) fields due to sources like current flow which drags A along with it and unreal A = del chi (and I question that it ever exists or even really has heuristic use). Some background A fields may not count in lab situations, like the large rotating A field of the earth passing uniformly through a relatively small lab (+ constant A) which wont affect lab tests. Lorenz gauge

is a tradeoff between space and time for propagating waves (and I dont understand that well).
2) 'Real A can give real B = curl A and real E = -dA/dt. I count AB (Aharonov-Bohm A) because it can be seen in a lab. The A that can give B or AB can be derived from a phase field over space using directional derivatives. We could add a del chi (if that could actually be done) or constant A field onto real A without changing B or AB. AB only sees phase differences between paths rather than phase itself.
3) Real sources like current flow cause real A which makes phase vary locally through space (but in a controlled way− not random or arbitrary). Phase change has a cause. The interference patterns exist because the cheat term eA does act like EM momentum changing phases differently on one side of a solenoid versus the other. That is, there are irrelevant phase changes (most of which I don't think exist but yet are stressed in gauge transformations) and there are relevant phase changes that really do affect electrons. We just don't use those in del chi.
4) I think the usual presentation of gauge principle is intrinsically confusing and only inadvertently guides the form of something than can be real (the A 'buddy' of del chi). I really believe the whole thing could be done better (and some examples would help, and the best example is the interior solenoid with constant B field).
5) The gauge principle gives hints about the form of SU(2), SU(3) fields. But we know EM pretty well already and dont need the hint, we can work backwards to see what is really going on. Modify the psi from the SCH equation without A by phase changes to get the SCH equation with A (the EM SCH eqn). They say A = A + del chi (but the A in that case was 0, and its real nature is never discussed). Their A is a ghostly cheat. But it doesnt have to be that way.
6) The phase change on psi is another solution $[exp(iq\varphi/\hbar)]$ times ψ_o where the phi is a phase integral $\int A \cdot dx$ (Aitchison). That works well for both the directional derivative A and the gradient (del chi) which is also a directional derivative but in the maximal change direction. I think this is the key of the problem: both A and del chi can be processed in the phase integral. In that sense, a real $A' = A + \nabla\chi$ makes some sense together as buddies.

7) The gauge principle starts with a position of total ignorance about phases and postulates 'arbitrary freedom to change electron phase smoothly but locally as a single valued function over space-time. An example is a field of temperature values from which a gradient of temperature field can be formed. Associated with phase changes is a vector field which could be like del chi (artificial unreal useless gradient field). We focus on the useless del chi term and simply say that along with it could be an associated A field which could be useful. We do that because it works, and Yang-Mills extrapolated it to $SU(N)$'s. Useful A vector fields, in contrast, are associated with restricted classes of phase changes and also with causes like currents which give them directions. Although the gauge principle talks about 'arbitrary changes, the restriction to 'function isnt arbitrary enough for real physical cases (which can be multivalued like an advancing helix).

7. Is there a Genuine Vector Potential, $\vec{A}$?

Despite experimental verification of the vector Aharonov-Bohm effect (AB), many still doubt the "reality" of the vector potential because of its non-uniqueness: it is subject to a choice of a Gauge Convention such as the Coulomb gauge, $\nabla \cdot \vec{A} = 0$. Is it still possible to argue for the physical existence of a basic central core vector potential? Does Nature utilize a unique single choice for $\vec{A}$? Suppose one **defined** such a classical $\vec{A}$ field as one due to known electrical currents $\vec{J}$ which can result in either a non-zero magnetic field $\vec{B} = \nabla \times \vec{A}$ or a non-zero AB phase shift.
That is, if Ampere's Law has the form:

$$\nabla \times \vec{B} = \mu_o \vec{J} = \nabla \times \nabla \times \vec{A} = \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}. \tag{37}$$

Assume the Coulomb gauge $\nabla \cdot \vec{A} = 0$ above so that we are left with the vector Poisson equation:

$$\nabla^2 \vec{A} = -\mu_o \vec{J}, \quad \Rightarrow \vec{A} = \frac{\mu_o}{4\pi} \int \frac{\vec{J} d^3 x}{r} \tag{38}$$

This form is a special case of the Lienard-Wiechert integral for the potential. So, the current sources determine the vector potential subject to Coulomb gauge condition. Note that this is a very natural and desirable condition because we don't wish our A-field to look like an E-field derived from Gauss' Law ($\nabla \cdot \vec{E} = \rho/\epsilon_o$).

As examples, consider again the previous cases of the interior (in) and exterior (out) fields of an ideal solenoid.

$$\vec{A}_{in} = \mu_o I n \rho \hat{\phi}/2 = B_o \rho \hat{\phi}/2; \; \vec{A}_{out} = \vec{A}_\phi = K \hat{\phi}/\rho; \quad [K = \mu_o n I R^2/2 = B_o R^2/2]. \tag{39}$$

where the interior magnetic field is simply $B_o = \mu_o I n$ − current times density of turns. These results also satisfy the Coulomb gauge condition because there is no variable phi dependence. In cylindrical coordinates (ρ, ϕ, z), the divergence operator is given by:

$$CYL: \quad \nabla \cdot \vec{A} = \frac{1}{\rho}\left[\frac{\partial}{\partial \rho}(\rho A_\rho) + \frac{\partial}{\partial \phi} A_\phi + \frac{\partial}{\partial z}(\rho A_z)\right], \tag{40}$$

and, in our case, $(\partial A_\phi / \partial \phi) = 0$.

Suppose we had worked backwards from a known solenoid magnetic field to deduce the possible A field using curl A. Then, for the $B = 0$ case we have:

$$(\nabla \times \vec{A}_{out})_z = \frac{\hat{z}}{\rho}\left[\frac{\partial(\rho A_\phi)}{\partial \rho} - \frac{\partial A_\rho}{\partial \phi}\right] = 0 = (\hat{z}/\rho)[\rho \partial A_\phi/\partial \rho + A_\phi - \partial A_\rho/\partial \phi]. \tag{41}$$

Try a simple solution for A of the form: $A_\phi = a\rho^b \phi^c$ and also include $A_\rho = \alpha \rho^\beta \phi^\gamma$ and set the above term in brackets to zero. Then we obtain in general

$$(A_\phi, \; A_\rho) = (a\rho^b \phi^c, \; \frac{a(b+1)}{c+1} \rho^b \phi^{c+1}). \tag{42}$$

The usual outer solenoid vector potential solution $A_{\phi(out)} = K/\rho$ corresponds to $a = K, b = -1, c = 0$; so $(b+1) = 0$ and there is no radial term. If the inner potential had derived from a gradient field (which it does not!), then we would have $b = +1, c = 0$ giving: $(A_\phi,\, A_\rho) = (B\rho)[\hat{\phi}/2 + \hat{\rho}\phi]$. And this would have given curl $A = 0$; and the actual solution lacks the radial term and yields $B = B_o > 0$.
The point is that we could have added functions such as $(a\rho, 2a\rho\phi)$, or $(a, a\phi)$, or $(a\rho^2\phi^2, a\rho^2\phi^3)$ onto any solution and have these terms also contribute zero magnetic field in the curl. But terms such as these come from no obvious source, and the increase with angle phi makes no sense. There is the previous solution for $A_\phi(\phi, \rho)$ that does make sense, has clear source, and can be considered as a 'core A field.'

Another reason for belief in a 'core A' is envisioned in the possible way that Nature may calculate it. We would like to think of A as analog to the Lense-Thirring (1918) 'dragging of inertial frames' in general relativity. For example, a rotating mass drags space-time with it via metric terms h_{oj} that are similar to the electromagnetic vector potential, A. Calculation of 'dragging' can be done in a manner resembling cellular-automata.

As an example of space-time computation, "How do fields propagate?" A standard numerical technique for solving Laplaces equation, $\nabla^2\phi = 0$, is to divide up the space between two potential boundary surfaces into cells and iteratively update the value of the potential ϕ in each cell over the average of the values in its neighboring cells until numerical results are stable $-$ using some variant of Jacobi or Gauss-Seidel iterations. For example an $n+1$'th iteration from previous n'th values at 3D x,y,z cell i, j, k may look like:

$$\phi{*}_{i,j,k} = (\phi_{i+1,j,k} + \phi_{i-1,j,k} + \phi_{i,j+1,k} + \phi_{i,j-1,k} + \phi_{i,j,k+1} + \phi_{i,j,k-1})/6. \tag{43}$$

So, over many successive iterations, the values of the cells at the boundary condition propagate to the free space producing an array of self-consistent values. It is highly tempting to say that this is how Nature also does it - **simple average with respect to neighboring space-time 'cells'** $-$ and does so for many different type of field. One might say that each region of space continually calculates and stores and updates and processes these values. Note that cellular atomata (CA) update cell states in terms of states of neighboring cells. The idea above of a regular mesh for spacetime is just a visual convenience; and the idea that each cell carries a real-number potential may also be a stretch. Perhaps this is more like a phase value instead $-$ the modulo 2π of larger range numbers. Then neighboring cells would compare the differing values of the phases as key information.

Laplace's equation applies to electrostatic problems in free space, magnetostatic problems, the Newtonian gravitational potential and can be made to apply to the vector Laplacian and even to tensor Laplacians such as the weak gravitational metric, h_{ij}. In the case of the vector Laplacian for free space, we can write:
$\nabla^2\vec{A} = 0 = (\hat{i}\nabla^2 A_x,\ \hat{j}\nabla^2 A_y,\ \hat{k}\nabla^2 A_z) = (0,0,0)$ for the equivalent of three separate

Laplace's equations [10] . Each can be solved by the iterative method. Thus, the iterative average concept can have fairly wide application implying again that it may faithfully represent reality. Just because we cannot see a unique $\vec{A}$ doesn't mean that Nature doesn't use one.

References

[1] Bruce A. Schumm, Deep Down Things, The Breathtaking Beauty of Particle Physics, Johns Hopkins Press, 2004.
[2] Y. Aharonov and D. Bohm, "Significance of Electromagnetic Potentials in the Quantum Theory," The Physical Review 115 No. 3, August 1, 1959, pp. 485-491.
[3] Herbert Goldstein, Classical Mechanics, Addison-Wesley Publishing, 1950.
[4] I.J.R. Aitchison and A.J.G. Hey, Gauge Theories in Particle Physics, Second Edition, Adam Hilger, 1989.
[5] Bruce J. Hunt, "Oliver Heaviside, A first-rate oddity," Physics Today, Vol .65, No. 11, Nov 2012 pp 48-54.
[6] A.O. Barut, Electrodynamics and Classical Theory of Fields and Particles, Dover Publications, 1964, 1980.
[7] D.L. Peterson, "Einstein's formula for Mach's Principle," "The Lense-Thirring effect," "An analogy between the linearized theory of general relativity and electromagnetism," unpublished reports, 1976-1978. And AB description, 10/2/85.
[8] Paul H. Frampton, Gauge Field Theories, 2008. Book section on the web, Gauge Invariance http://www.wiley-vch.de/books/sample/3527408355_c01.pdf
[9] Wikipedia articles on gauge term definitions.
[10] Francis Halzen, Alan D. Martin, Quarks & Leptons: An introductory course in modern particle physics, John Wiley, 1984.
[11] J.D. Jackson, "Historical roots of gauge invariance," Rev. Modern Physics, Vol 73, July 2001, p6630-672.
[12] Theodore Frankel, The Geometry of Physics, Cambridge, 1997, 2004.
[13] K. Moriyasu, An Elementary Primer for Gauge Theory, World Scientific, Singapore, 1983.
[14] Gregory L. Naber, "Gauge Fields in Physics and Mathematics," www.pages.drexel.edu /∼gln22/GaugeFields.pdf
[15] Roger Penrose, The Road to Reality - a complete guide to the laws of the universe, Knopf, 2005.
[16] www. Scholarpedia.org/article/Gauge_Theories
[17] David Peterson, "Circle Functions in Modern Physics," LaTeX report–unpublished–updated to March 6, 2012.
[18] Physics Today, "Gauge symmetry saved, mass endowed," September 2012 p 14.
[19] Watson Fulks, Advanced Calculus, Second Ed., John Wiley and Sons, N.Y., 1961, 1969. p. 343.
[20] David McMahon, Quantum Field Theory Demystified, McGraw Hill N.Y., 2008
[21] Robert Lleighton, Principles of Modern Physics, McGraw-Hill, 1959
[22] David Griffiths, Introduction to Elelmentary Particles, 2nd Ed., Ch 10: Gauge Theories. 2008.
[23] N.P. Konopleva, V.N. Popov, Gauge Fields, from Russian, 1981. QC793.3
[24] Claude Cohen-Tannoudji, Bernard Diu, Franck Lalo e, Quantum Mechanics, Volume One, John Wiley and Sons, N.Y., 1977.

[10] But this simple decomposition only applies for the special case of Cartesian coordinates. Other curvilinear coordinate system cases are much more complex.

[25] C.N. Yang and R.L. Mills, "Conservation of Isotopic spin gauge invariance," Phys. Rev. 96 (1954), 191-195.

8. Appendix: **Definitions:**

Gauge: "Gauge" means a convention or standard regarding potentials or 'sub-potential' gauge functions but can also refer to wave-function phase. Gauge refers to redundant degrees of freedom in an appropriate Lagrangian, $\mathcal{L}$. The term 'gauging' means localizing or changing a global symmetry into a local symmetry along with introduced additional compensating fields. Older common usage was the setting of a standard of linear distance or a convention of dimension or scale. But, in modern physics use, it no longer means 'scale.' Gauge is a bad name, but we are now stuck with it.

Gauge Theory: refers to a quantum field theory using a Lagrangian which is invariant under a continuous group of local transformations. The set of gauge transformations between redundant gauges forms a continuous Lie symmetry group or gauge group. 'From a physicist's point of view, the existence of a symmetry implies that some quantity is unmeasurable' [10] − it has no effect on measurable physics. Associated with any Lie group is the Lie algebra of group generators. For each group generator there necessarily arises a corresponding vector field called the gauge field [9]. Gauge fields are included in the Lagrangian to ensure its invariance under the local group transformations (called gauge invariance). Or, restated, 'a vector field such as **A** introduced in order to guarantee local phase invariance, is called a gauge field.' [4]. A goal of compensating fields is to keep the symmetry unobservable. When a gauge theory is quantized, the quanta of the gauge fields are called gauge bosons. If the symmetry group is non-commutative, the gauge theory is referred to as non-Abelian, the usual example being the Yang-Mills theory. The modern era of gauge theories began with the 1954 paper by Yang and Mills. Instead of 'Gauge Theory,' mathematicians prefer to use the 'Theory of Principle Fibre Bundles.'

Gauge Bosons: presently refer to the quanta of the interacting gauge fields such as the twelve gauge bosons: the photon, three weak bosons and eight gluons of the "Standard Model" represented by the symmetry group $U(1)_Y \times SU(2)_L \times SU(3)_C$. But this concept could be extended to supersymmetry and SO(10). For each Lie group, there are as many gauge bosons as there are generators of the gauge field. For $SU(2)\times U(1)$, the gauge bosons (W^1, W^2, W^3, B^o) get rearranged by spontaneous symmetry breaking and the Higgs field and become $(W^+, W^-, Z^o, and \gamma)$ gauge bosons while also gaining mass for the weak boson set. The photon γ is represented by the A-field. Color is unaffected by Higgs, so gluons stay unbroken the same as they were before.

Gauge Convention: such as the Coulomb gauge, $\nabla \cdot \vec{A} = 0$, or the Lorenz gauge condition so essential to EM wave equations, [11] $\nabla \cdot \vec{A} + \epsilon\mu\partial V/\partial t = 0$, means a gauge

[11] also called choosing a gauge and named after Ludvig Lorenz NOT Hendrick Lorentz − an error perpetuated through many textbooks.

standard or gauge selection or 'gauge fixing' which restricts or constrains the functional freedom of electromagnetic field potentials. This, of course, has nothing to do with lengths but merely represents a useful conventional form or standard 'constraint' over a sum of potential values. This concept also applies to classical field theories. This gauge condition is a mathematically convenient supplementary equation which helps to suppress the indeterminacy of the potentials. In some cases, the gauge condition reveals a coupling between scalar and vector potentials.

Gauge Function: is the physically irrelevant 'sub-potential' that can be added onto EM potentials in the form of $\partial_\mu \chi(x,t)$. It might be a purely mathematical addition without any actual reality in Nature. But $\chi(x,t)$ does introduce a phase shift, and that makes one wonder. χ is a scalar field which couples to a change in QM phase. This hidden and ghostly (or 'chindi') chi field is a scalar field attaching some smoothly changing value to each point of space-time (e.g., a temperature field over 3-space). Unlike the core A field, this fields values and gradients are path independent– like a topo-map whose heights are approachable from any direction. But these hills can change with time. Gauge transformations refer to transformations of potentials using the in-essential and irrelevant chi fields, changing something that is unimportant $-$ and changes in QM phase. For SU(2) there are two of these functions (χ_1, χ_2), and for SU(3) there are three phases for $\vec{\chi} = (\chi_1, \chi_2, \chi_3)$.

Gauge Principle: is a stated procedure for obtaining physical interaction terms from a free Lagrangian which is symmetric with respect to a continuous symmetry (e.g., U(1) phase symmetry). The gauge principle provides a method to transform a Lagrangian which is invariant under a global symmetry into a Lagrangian that is invariant under a local symmetry (gauge invariant). For $U(1)$, it involves going from ordinary momentum operator derivatives to covariant derivatives which include a gauge field, A_μ, but also adding a kinetic energy term containing the strength tensor $F_{\mu\nu}$ (a 4-curl of A). Relevant physics is required to be invariant under the local gauge transformation $\psi' = e^{iq\chi/\hbar}\,\psi$, and this determines the transformations and form required for the potentials A_μ. Gauge Invariance or Gauge Symmetry means the invariance of a theory under the combined transformation like ψ', A', ϕ'. The intent [Fock 1926, Weyl 1929] is that the existence of 4-vector potentials (and field strengths) follow from the phase invariance of matter fields [11]. Fock used the term "gradient invariance" $-$ which in turn says something about the importance of curls (or generalized curls such as $F_{\mu\nu}$). Noethers Second Theorem says that local Lagrangian symmetries imply new gauge fields.

Gauge Covariant Derivative: is the ordinary derivative operator modified with additional terms enabling a preservation of needed physical properties under gauge transformations. It is like a generalization of the covariant derivative used in general relativity which adds Christoffel symbol terms (${\Gamma^i}_{jk}$) for parallelism under curvatures (in particular for 'affine connections'). An example for quantum field theory is $D_\mu = \partial_\mu - ieA_\mu$. The vector potential A is sometimes called the 'gauge potential' [13] or the 'gauge connection.'

A change of phase in the wavefunction, ψ has to be accompanied by a change in A. The Christoffel symbol for 5-D Kaluza theory is $\Gamma_{5\mu,\nu} = (\kappa/2)F_{\mu\nu}$. [12] But the goal of the connection in curved manifolds is the parallel transport of vectors. In gauge theory, the goal of the connection potential is the translation of phase factors over paths. [14]

Cheating Terms in the modified derivative Covariant derivatives such as $D_\mu \equiv \partial_\mu - ieA_\mu$ introduce 'compensating gauge fields' like A_μ also called "cheating terms." QED has a U(1) local gauge invariance or local phase transformations. *Compensation prevents observability of unmeasurables*, and absolute quantum phase is unmeasurable. Gluon fields are needed to compensate phase changes for QCD, and weak fields compensate phase changes for electroweak theory. The gauge field is a non-integrable phase factor (Dirac 1931, AB 1959).

Canonical Momentum: is the generalized momentum conjugate to position, $p_j = \partial L/\partial \dot{q}_j$. For EM fields, $p = mv + qA$ where mv is called the usual 'mechanical momentum' [3]. The symbol π is sometimes used for momentum density for continuous systems, $\pi = \partial \mathcal{L}/\partial \dot{q} -$ e.g., $\mathcal{L} = (E^2 - B^2)/2 - \rho\phi + j \cdot A$. The jA term is an 'interaction' between a Noether current, j, and the A-field.

Gauge group: expresses the gauge symmetry invariance under gauge transformations. A gauge transformation such as $A_\mu \to A_\mu + \partial\chi/\partial x_\mu$ maintains an equivalent Lagrangian. The Lie groups are the most interesting continuous groups (Sophus Lie, 1842-1899). The Yang-Mills isotopic spin local symmetry gauge group was SU(2). Other examples are electromagnetism U(1), weak $SU(2) \times U(1)$, and color SU(3). Mathematicians use the term 'structure group.'

Yang-Mills theory: [Chen Ning Yang and Robert Mills, 1954, 'YM' [25]] is a non-abelian gauge theory based on the $SU(N)$ group and intended as a generalization of Maxwells $U(1)$ electromagnetism. Yang and Mills invented a new non-abelian field strength formula largely by trial and error experimentation with an initial focus on Heisenbergs isotopic spin and allowing independent local changes in isotopic spin direction. The model was initially considered as unreal because it required massless particles for preserving gauge invariance while pions or weak bosons are massive particles. A 'Pure YM' theory has only a gauge field without an associated matter field. So, for example, quantum $SU(3)$ YM theory describes gluons in the absence of quarks. QCD gauge field theory is obtained from $SU(3)$ YM theory by coupling it to fermionic quark fields.
The point of gauge symmetry is that it constrains the form of the action and dictates the form of the interaction. This was Yang's goal using the principle of local symmetry. The application of YM to HEP was very tricky in requiring a developing understanding of Higgs

[12] In GRT, the Kaluza 5-D metric contains a row and column of A's, and the GRT 'connection' is formed from derivatives of this metric − so the Christoffel symbol uses E and B fields components rather than the different gauge or fiber connection, A.

symmetry breaking and quark confinement.

Renormalization is any of a collection of techniques used to treat infinities arising in calculated quantities. In quantum field theory, it is a procedure " by which divergent parts of a calculation, leading to nonsensical infinite results, are absorbed by redefinition into a few measurable quantities, so yielding finite answers." Quantities such as mass and charge in the mathematics of a theory may not correspond to those actually measured in a laboratory. Bare quantities can be redefined to measured quantities. A major breakthrough in the infinites that plague QFT was the 1971 paper by 't Hooft showing that renormalizable theories must be Yang-Mills local gauge theories. The Higgs model allows the generation of masses for weak bosons without spoiling the renormalizability of EW gauge theory. Gauge theory arranges systematic cancellation of divergences between pairs of Feynman diagrams.

In simpler terms, charge renormalization is nothing more than vacuum polarization (screening). The usefulness of the Higgs mechanism results from the Higgs field condensate being a perfect dielectric. In non-abelian gauge theory charge renormalization is 'anti-screening' from gluons acting as color magnetic dipoles causing YM vacuum paramagnetism. QCD anti-screening overcomes quark screening by a factor of $+11$ to -2.

In the Language of **Fibre Bundles:** Maxwell's EM field F can be represented using a principal bundle of group G = U(1). A connection on the bundle is the EM potential, and its curvature is the EM field. The connection will depend on the gauge, but the curvature is gauge invariant. "Gauge transformations are simply changes of frames in the fibers of the bundle" [12]. Yang and Mills had been unaware of the concept of curvature of a vector bundle. A gauge field is a connection (ω or A), i.e., $\omega = -iqA$, and the curvature is $F = i\theta/q$ is also called the field strength: $F_{jk} \equiv \partial_j A_k - \partial_k A_j - iq[A_j, A_k]$. Curvature is a 2-form $\theta \equiv d\omega + \omega \wedge \omega$ or $F = dA + A \wedge A$, where $\wedge$ is 'wedge product.' The Y-M equation is the Hilbert space adjoint $\nabla^*\theta = 0$.
The term "bundle" means a differential manifold consisting of a total space B, a base space M, and a projection map $\pi : B \to M$. If the fiber is a Lie group G, then there is a principle bundle $P = (P, M, \pi, G)$.

<u>Comment: Derivation from inserting A values into a fifth dimension in a general relativity</u> 5×5 metric tensor: This was Kaluza's approach, and it did furnish the source-free restricted Maxwell field equations (no media and no charge/current sources). Kaluza obtained the classical homogeneous vacuum solutions in 1919 (the relativistic version $\partial_{[\mu} F_{\nu\lambda]} = 0$ of Faraday's Law and $\nabla \cdot B = 0$). Klein interpreted the 5th dimension as little U(1) circles. Kaluza-Klein ideas (KK) have now revived using multiple hidden dimensions (e.g., supergravity). KK fails to give the inhomogenious Maxwell equations in part because source currents were never needed since A was placed into the metric tensor by hand.

Five Dimensional view of Electromagnetism

Dave Peterson, 9/15/05
Revised: 5/18/07-10/17/07

In 1919, Theodor Kaluza submitted a five-dimensional or "5-D" version of unified gravitation and electromagnetism to Albert Einstein. Kaluza added an extra row and column onto Einstein's 4x4 metric tensor $\{g_{\mu\nu}\}$ and inserted the EM vector potential A_μ and a scalar term into those locations. The Kaluza miracle was that the formation of equations of motion turned into those of a combined gravitation and Maxwell EM field in 4-dimensions. Einstein had alternating mixed feelings about the work but allowed it to be published two years later. There was an immediate problem that no one had detected a 5^{th} dimension. So Oskar Klein suggested that it is curled up into a tiny circle so small that it cannot be directly detected. This produces Maxwell's equations for electro-magnetism and quantized charge -- except that the e/m ratio is unreal and requires additional theory. In some sense, E&M is a metric theory using each tiny circle attached to every point of space-time. Using input from both these authors gives "Kaluza-Klein" theory or "KK."

If this interpretation of electromagnetism gives the correct relativistic electrodynamics, why isn't it taught in colleges? One answer is that although it works and gives relativistic E&M, it presently doesn't predict anything new. Science works by making predictions and showing that they are experimentally true or false. There is no present ability to perform an experiment actually testing "KK" and the real existence of a fifth dimension. Actually, by the 1980's, KK theory had largely been forgotten and was even hard to find in most books on gravitation.

However, in current times, this approach to unified field theory has sparked a whole industry with up to 7 "Kaluza-Klein" dimensions plus the usual 4-D space time. The straight "KK" idea in 11 dimensions is called supergravity. But, with several recent "super-string revolutions," it is now realized that the KK approach is somewhat good but not quite perfect and may just be a weak field approximation of something much more complex {the mysterious not and quite developed "M" theory}. Although many physicists believe in super-strings, they also suffer from the problem that it is very difficult or even unlikely that super-strings can be tested as a scientific theory.

In publications between 1861 and 1873, Maxwell had an inspiration that a changing electrical current should result in a magnetic field. Combining this with Faraday's observation from 1831 that a changing magnetic field could produce an electric field, Maxwell calculated that electromagnetic radiation should exist and should move with the speed of light, c. The subsequent experimental discovery of electromagnetic radiation was a victory for Maxwell's equations.
The overall conceptual foundation of electro-magnetism has changed since Maxwell's days. E&M was recognized to be a relativistic theory already containing Lorentz transformations. Einstein's relativity paper of 1905 was actually titled, "On the Electrodynamics of Moving Bodies" because it pulled relativity out of E&M. Rather than separate electric fields, E, and magnetic fields, B, there is really a single invariant entity called the **anti-symmetric tensor $F_{\mu\nu}$.** For the early half of the 20^{th} century, most people

believed that only the fields themselves had existence because they had energy-equivalence. It was believed that the electromagnetic potentials A and ϕ were only mathematical conveniences for calculation. But Yakir Aharonov and David Bohm showed that the potentials were quite real. In particular, they can change the location of electron interference fringes even in the absence of any E or B fields in the path of the electron. Modern quantum mechanics writes an effective electron momentum as $\pi = p + eA$ as if the vector potential itself represented a flow of something and eA is some sort of "electromagnetic momentum." The electron may be thought of as having a charge part and a mass part with inertial momentum depending on mass and electromagnetic momentum associated with charge and vector potential. The Lorentz term $F = qv \times B$ might be thought of as a Coriolis like effect due to being in the wrong frame of reference. It does look strongly like the real Coriolis effect for mass $F = 2m\,\omega \times v$, where B takes the place of omega – the "swirl of the A field."

The vector potential is derived from electric currents. In MKSA(SI units), $B = \nabla x A$, and

$$A(r_2) = \frac{\mu_o}{4\pi}\int\frac{J(r_1)}{|r_2 - r_1|}\,dV_1 = \frac{\mu_o I}{4\pi}\oint\frac{dr_1}{|r_2 - r_1|}\;;\; E = -\nabla\phi - \frac{\partial A}{\partial t},\; and\; \nabla x E = -\frac{\partial B}{\partial t}.$$

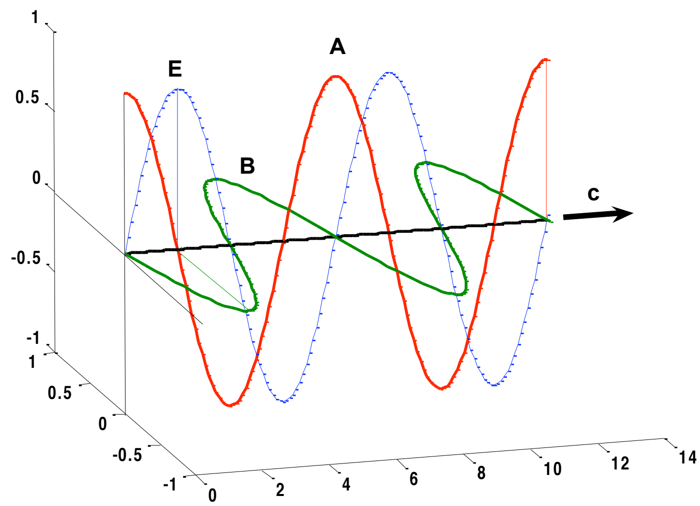


***Figure 1**: Crossed E and B fields can also be pictured as just an **A** vector-potential field. This picture has a plane polarized Electric field. The argument ($kx - \omega t$) here uses t = 0 and k = 1.].

Traveling radio waves are sometimes pictured as crossed E and B fields moving at the speed of light. If a radio tower has current flowing quickly up and down the tower, it is also possible to think of the resulting waves as just an **A** field also moving up and down in space and propagating outwards from the tower. Then $E = -\,\partial A/\partial t$ and B = curl of A are "derived fields" from A. Whether it is strictly true or not, **A** can be pictured as the electric current dragging some sort of adjoining electro-magnetic space along with it.

But what about waves with circular polarization? That case can be thought of as being due to two sets of currents perpendicular to each other and out of phase with each other so that the A vector has a helical profile propagating through space (e.g., an "extra-ordinary" current parallel to an optic axis and also an "ordinary" current perpendicular to an optic axis in a crystal). Then $E = -\partial(A_o + A_{eo})/\partial t$.

For electric current moving down a long wire in the "z" direction, the vector potential A will drop off with distance as $A_z = -\ln|r|\,(\mu I/2\pi) + A_o$. An arbitrary background constant field could exist and will not affect E&M physical results.
Electromagnetic induction may be conceived of as: an accelerating charge at one point in space produces a time changing A field at another charge some distance away, and Lenz's law results in an induced Electric field E = - dA/dt which can move that charge.

In the "KK" picture, this dragging might really be the scalar field of phases of little circles moving clockwise or counter-clockwise at each point of space-time. In the simplest view, attaching a little circle to each point of space-time may temp one to attach a single scalar phase to each circle. The problem with that is that the vector potential may be initially thought of as a gradient of that scalar field, but the curl of a gradient is zero. That means that no magnetic field can be derived from such a simplistic picture. General relativity is a "metric" theory. That means that a separate coefficient can be attached to each cell of the metric tensor, g. For static fields, one usually has a non-zero coefficient in front of just the diagonal terms. But for dynamic cases like the rotation of a massive body, there can be separate functions for each off-diagonal term like dxdt, dydt, dzdt – three possibilities corresponding to the possibility of "frame dragging". Similarly, if the 5-th dimension is labeled by "ζ ," there can be new coefficients in front of the mixed differentials $dxd\zeta$, $dyd\zeta$, and $dzd\zeta$. These coefficients <u>are</u> the potential A_x, A_y, and A_z. And by analogy, the zeta dimension would almost appear to be some sort of electromagnetic "time."

In General Relativity, "Space-time tells matter how to move and matter tells space-time how to curve." "$G = 8\pi T$" where "G" is the "Einstein Tensor" for curvature and T is the "stress-energy tension" for sources (mass, momentum, pressure). KK equations do not show any sources for E&M. The results of E&M sources are already represented by the vector potential A_μ in the 5-D metric tensor $g_{\alpha\beta}$.

Kaluza "imposed a somewhat artificial restriction" called the "cylinder condition" barring the 5th coordinate from "making a direct appearance in the laws of physics [1]." He then obtained the electromagnetic photon from an empty 5-dimensional space-time:

{$R_{AB} = 0$ or $G_{AB} = 0$ in 5-D implies that ${}^4G_{\mu\nu} = {}^4T_{\mu\nu}{}^{EM}$ in 4-D – The "KK" Miracle}.
T is the energy-momentum tensor which now represents sources where none existed before.

This is also the key concept motivating higher dimensional unification. So EM can arise purely from the geometry of empty space, and that was "Einstein's vision" and long-sought goal.

In ordinary special relativity, physics often "appears" to be 3-D because of the largeness of the speed of light, c. Most of our world has slow motions, v << c. The true mixing of space and time appears only at very high speeds. For "KK", A_μ in the metric tensor is not seen directly because of the tiny compactification of the coordinate x_5 or the insistence that all derivatives be zero, $\partial\ (\ )/\partial x_5 \equiv 0$.

In quantum mechanics, EM also results from imposing a one-dimensional group U(1) gauge-invariance on a free particle Lagrangian. This is equivalent to “invariance with respect to coordinate transformations along the fifth dimension [1].” Klein came to the conclusion that the size of the 5th dimension should be truly tiny – comparable to the Planck distance ~ 10^{-35} m. Klein compactification gives the appearance of explaining charge quantification – but this claim was later abandoned because the e/m ratio seemed to be wildly inappropriate (e.g., m ~ m_{Planck} ~ 10^{19} GeV instead of 0.5 MeV). This observation is also a main reason that most people abandoned 5-D KK theory. The non-observability of the 5th dimension and the introduction of a scalar dilation field were also troublesome – but scalar fields are much more tolerable in the present time.

“SuperGravity” (SG) is largely a KK theory in 11 dimensions. But both supersymmetry (“SS”) and the new unified M-Theory now yield SG only as a weak field approximation. “Kaluza’s original aim of explaining forces in geometrical terms is thus abandoned completely.” (in strict principle).
Ed Witten [3] says that the Kaluza-Klein theory successfully “unifies the metric tensor g and a gauge field A into the unified structure of five-dimensional general relativity.” “While the Kaluza-Klein approach has always been one of the most intriguing ideas concerning unification of gauge fields with general relativity, it has languished because of the absence of a realistic model with distinctive testable predictions.”
Reality is probably more complex than the extensions of the Kaluza-Klein approach, but it is still semi-real and constructive. The ultimate extension is close to eleven-dimensional supergravity.
It is not perfectly clear if 5-dimensional Kaluza-Klein theory should be formed from a five dimensional Minkowski space, M^5 , or as M^4 X S^1. The S^1 circle is more interesting because of the U(1) group of rotations of the circle and its relation to electromagnetism.

One form the the 5-D metric tensor would be:

$$g_{AB} = \frac{1}{\phi^{1/3}}\begin{pmatrix} g_{AB} + \kappa^2\phi A_\alpha A_\beta & \kappa\phi A_\alpha \\ \kappa\phi A_\beta & \phi \end{pmatrix}$$

which has a side row and column containing A_μ. This looks like the g_{oi} or g_{io} general relativity case for “gravito-electromagnetism” (the dragging of space-time about a spinning body). There has always been a similarity between E&M and first order approximations to GRT. Perhaps placing A_i in place of g_{oi} in the 5th dimension cells of $g_{\alpha\beta}$ is similar to treating A as some sort of frame dragging also. But then, one might also think of ζ as some sort of separate electromagnetic time.

The fields g, A, ϕ in the modern parlance of Quantum Field Theory would be the spin-2 graviton, the spin-1 photon, and the spin-0 dilaton. “The masslessness of the graviton is due to general covariance, the masslessness of the photon to gauge invariance, and the dilaton is massless because it is the Goldston boson associated with the spontaneous breakdown of the global scale invariance [2].” In some modern theories, elementary particles may be viewed as black holes. These are called “extreme black holes” and are not subject to Hawking radiation.”

Gauge Invariance:

Magnetic field B = ∇ x A = ∇ x (A' + ∇χ) = ∇ x A' + 0 (since the curl of the gradient is zero). So, the vector potential is arbitrary up to an additive factor of "del chi" ∇χ (also called the "Coulomb gauge"). The word "gauge" is an old term going back to Weyl where changes in scale were being considered. It really isn't quite appropriate anymore and might be replaced by some other word such as "phase." For relativistic 4-vector notation, $A_\mu \rightarrow A_\mu + \partial_\mu \chi$.

Vectors can be derived from gradients. But the useful vector potential A cannot itself be derived from a gradient because then there would be no resulting magnetic field, B = ∇xA.

Remember that Maxwell's equations resulted from the 5-D metric for both the Klein approach with a compactified 5-th Dimension and the Kaluza approach with a non-compactified dimension. The inputs to the 5-D metric tensor include the full A vector field, and a vector field is much richer than a scalar field and its gradients. One view is that the A vector momentum could be replaced by wave-numbers of increasing phases in the x, y, and z directions together – not just one direction. So **A** is like **k** as a vector. These little Klein circles are not really separate discrete circles because they are continuous with space-time. For example, S^1 x R = circle x line = a garden hose—or more precisely like a helix progressing along a dimension. It is possible to think of crossing a circle with a line in the x direction, a line in the y direction, and a line in the z direction all separately. The phases can be tracked separately in all three directions. Some may progress rapidly and some slowly. If the phases are "chi", then there is a phase χ_x, χ_y, and χ_z.. Other subscripts could be used for other coordinate systems (e.g., cylindrical ρ, z, ϕ.

For the case of a long current wire in the z direction, it may be possible to assign $A_z(\rho) = \partial\chi_\rho(r)/\partial z$ – without the other derivatives. Then, the vector potential A can be pictured as a succession of tiny circles with progressing phase moving up in the z direction.

As another case, consider the uniform B field between two Helmholz coils (use cylindrical coordinates, z,ρ, ϕ).

$$\vec{A} = \frac{B\rho\,\hat{\phi}}{2},\ so\ \vec{B} = B_o\hat{k} = \hat{k}\ \nabla x\vec{A}.$$

Then the vector potential A can be considered as little circles with phase increasing only in the angular ϕ direction. E.g., $A_\phi(\rho) = \partial\chi_\phi(\rho)/\partial\phi$.

The new phase picture from the Klein approach might be crudely thought of as looking at the valve of a bicycle tire as the tire moves around a circular hoop at radius ρ (the phase of the bicycle tire). For the long current wire, look at the valve of a bicycle tire as the tire moves on a long rod in the z direction at radius ρ from the current carrying wire.

The connection between these 3-D phases (χ_ρ, χ_z, χ_ϕ) and the phase of the quantum mechanical U(1) group is not perfectly clear.

This view of E&M as due to tiny compactified 5th dimension is pleasing and possibly true but presently untestable.

1. Overduin and Wesson, "Kaluza-Klein Gravity," arXiv:gr-qc/9805018 v1, 7 May 1998.
2. M.J. Duff, "Kaluza-Klein Theory in Perspective," arXiv:hep-th/9410046 v1 7 Oct 94.
3. Ed Witten, "Search for a realistic Kaluza-Klein theory, " Nuclear Physics B 186 (1981) 412-428.
* Note: [MATLAB should have used a right handed coordinate system.
4. Quarks and Leptons from Orbifolded Superstrings, Choi & Kim, Springer, '06.

Other Notes:
[4]. "But Kaluza's view of $g_{\mu\alpha}$ and $A^{\alpha}{}_{\mu}$ on the same footing has failed in fact because of the chirality problem."

[5]: The Trouble with Physics, Lee Smolin, 2006: Update: 11/1/06:
[pg 46]: "Nordstrom had found gravity by applying Maxwell's theory of electromagnetism to a five-dimensional world. Kaluza did this in reverse: He applied Einstein's general theory of relativity to a five dimensional world and found electromagnetism. You can visualize this new space by attaching a little circle to each point of ordinary three dimensional space." "If this theory is right, the electromagnetic field is just another name for the geometry of the fifth dimension."
"To get electromagnetism out of the [KK] theory, the radius of the circle must be frozen, changing in neither space nor time. This is the Achilles' heel of the whole enterprise and led directly to its failure. The reason is that freezing the radius of the extra dimension undermines the very essence of Einstein's theory of general relativity, which is that geometry is dynamical." A variable radius is unstable and could shrink to a singularity or grow to our size.

It is surprising to see Maxwell's equations spring from KK theory. Other derivations from Lagrangeans depend not only on A but also on energy terms like $E^2 - B^2$ as well. But perhaps it isn't so surprising considering that ordinary GRT already has concepts like gravito-magnetic fields (Lense-Thirring effects from off-diagonal h_{io} and h_{oi} terms) and gravito-electric fields (gravitational attraction). That sort of shape is built into GRT and gets duplicated with the extra 5-th dimension and it's A, ϕ terms.

Addition 2010: dp

Kaluza-Klein Electromagnetism and Fiber Bundles:

What is Electro-Magnetism [EM]? What is the physical as opposed to mathematical mechanism for the functioning of Maxwell's Equations? One approach to EM Foundations is the Kaluza-Klein [KK] mechanism where real space-time is 5-dimensional, and the 5-th dimension is composed mainly from the vector potential A_μ. KK is known to be an incomplete explanation, but perhaps string theory will ultimately complete it and allow KK in 5 dimensions as an approximation for weak field GRT + EM. Previous personal foundational attempts to find physical clarity in KK literature was not successful [5]. The main problem was that having the 5th dimension be a one-dimensional fiber existing at each point of Einstein's M^4 space-time seemed like simply making a scalar field for EM (an EM phase field-- like having a temperature over a volume). That view is not adequate to visualizing the richness of A_μ and Maxwell's equations—any gradient of a phase field would be gauged away. But no source in literature seemed to resolve this confusion. An explanation seems to be in the nature and richness of the 'connection' for the fiber bundle or for a metric space. "Trautman was the first to relate five-dimensional KK theory with the structure of fiber bundles." There is also confusion about whether the gauge theory examples apply only for quantum mechanics using an internal dimension or for classical theories—and the original KK was classical. Most of the new discussions of KK are for its quantum mechanical applications such as in string theory.

In its simplest form, a '**bundle'** is a differentiable manifold consisting of a total space B, a base space M, and a projection map π: B → M. A trivial example is the product manifold cylinder B = C^2 = R^1 x S^1 – a line and a circle. π: C^2 → S^1 = M base space, and R is the 'fiber manifold.' An example of B which is not a product manifold is the Mobius strip again over S and also with fibre R. But it takes two open sets to cover B (e.g., 2-sides). A base space could also be M = S^1/Z_2 = P^1 = one dimensional projective space—the space of diameters of a circle without arrows attached ['$/Z_2$' means that the identity map +1 on S^1 is identified with -1, the antipodal map]. The common 'tangent bundle' T(M) is a vector fiber bundle which is the collection of all tangent vectors at all points of M. Any point, p, on the base space is a set of tangents, π^{-1}(p). The global projection π takes the vector back to its sitting point, p on M.

The following notation is also used: B = (E, M, π; F) or Bundle = (total space, base space, continuous surjection; standard fiber). [8]. A trivial bundle is (MxF, M, pr_1, F) where pr_1: MxF → M. If the fiber F is a Lie group, G, then there is a 'principle bundle' P = (P, M, π; G). The immediate concern is G = U(1). An example of higher order Lie fiber groups is G = SL(2,C) and SU(2) for Loop quantum gravity spin or area network connections.

Gauge gravity uses external symmetries while gauge theory uses internal symmetries. In GRT, the vector bundle is the tangent bundle of space-time, and gauge potentials are analogous to affine connections or Christoffel symbols. And, in GRT, parallel transport is path dependent. The affine connection is a path or curve dependent identification of the tangent spaces of different points. The 'covariant derivative' of a vector is: $\mathbf{a^i{}_{;k} = \partial a^i/\partial x^k + \Gamma^i{}_{kr}\, a^r}$. There is an analogy to the gauge or bundle connection covariant derivative $\mathbf{D = \nabla_a = \partial/\partial x^a - ieA_a}$ or a shorter $\partial_\mu \rightarrow D_\mu = \partial_\mu - A_\mu$. The covariant derivative is the 'horizontal lift' of vectors tangent to the base space.

Maxwell's EM field $F_{\mu\nu}$ can be represented using a principal bundle of group G = U(1) and a connection on the bundle. The connection corresponds to the EM potential, and its curvature to the EM field. 'The connection will depend on the gauge, but the curvature is gauge invariant.' The connection is the true fundamental field. A fiber-bundle over space-time looks locally like the Cartesian product between the space-time and a fiber-bundle manifold. Kaluza -Klein (KK) theory uses an S^1 bundle over M^4. Having 5-D Ricci flatness R_{ab} = 0 and U(1) symmetry gives Maxwell's equations.

KK is only able to furnish the source-free Maxwell field equations (no J or ρ sources—so it is a limited theory). KK also suggests huge Planck scale values for electron m/e —so something else is needed to complete the theory. And Penrose believes that there are also problems with the stability of the size of the 5^{th} dimension.

Using a gauge choice called the 'cylinder condition' and weak-field-limit, Kaluza obtained the classical homogeneous $F_{\mu\nu,\lambda} + F_{\nu\lambda,\mu} + F_{\lambda\mu,\nu} = 0$ vacuum solution. In 1926, Klein tried to append quantum mechanics with his publication, "Quantum theory and

Five-Dimensional Theory of Relativity." He considered the 5-th dimension to be like a periodic phase with a U(1) invariant metric. There are analogies to gauge invariance of A fields and their accompanying QM phase transformation: $\psi(x) \rightarrow e^{i\theta}\psi(x) = U\psi$. Gauge and QM phase transformations are coupled together via the EM Schrodinger equation using $D_\mu = \partial_\mu - ieA_\mu$. Gauge, phase, and covariant derivative all go together.

"The family of phase transformations $U(\theta) = e^{i\theta}$, where a single parameter θ may run continuously over real numbers, forms a unitary Abelian group known as the U(1) group." [9]. In QED, Noether's theory with U(1) invariance implies the existence of a conserved current which becomes $\partial_\mu j^\mu = 0$ with $j^\mu = -e(\psi\text{bar})\gamma^\mu\psi$.

Traditionally, KK was a metric theory. For dynamic cases like the rotation of a massive body, there can be separate functions for each off-diagonal term of the metric like dxdt, dydt, dzdt – three possibilities corresponding to the possibility of "frame dragging" or gravito-magnetism. This uses h_{io} ~ a gravity version of A_i. Similarly, if the 5-th dimension is labeled by "ζ ," there can be new coefficients in front of the mixed differentials $dxd\zeta$, $dyd\zeta$, and $dzd\zeta$. These coefficients are the potential A_x, A_y, and A_z. [5]. In initial simplicity, one might think of the little fibers or Klein circles existing at each point of M^4 as adding single functional values to each point of M^4. But that would only provide a scalar-function. That concept by itself is not useful or relevant. Vectors can be derived from gradients. But the useful vector potential A cannot itself be derived from a gradient because then there would be no resulting magnetic field, $B = \nabla x A$. So each fiber provides more than just a value or phase. So, what is the new additional concept that is needed?

The 5th dimension is just a dimension—I was counting on it to supply the knowledge of the type of field present in space-time (values on the fiber to encode the A field). That's not the way metric theories work. The "KK miracle" is that Ricci-flatness in 5-D implies that ${}^4G_{\mu\nu} = {}^4T_{\mu\nu}^{EM}$ in 4-D spacetime—sources where none existed before. But T tells G how to curve—it is the curvature or connections that supply the knowledge of the A field and its derivatives $F_{\mu\nu}$. The Riemann tensor supplies values for path dependent changes in vectors traversing differing transplantations. Penrose says, "The curvature of our bundle connection then turns out to be the Maxwell field tensor F_{ab}." [10]. And Kaku says, $\boldsymbol{\Gamma_{5\mu,\nu} = (1/2)\kappa F_{\mu\nu}}$. **!!** Also $R_{5\nu} = -\alpha\partial^\mu F_{\mu\nu}$.

Unfortunately, it is hard to picture the curvatures—which was the original goal. So, the problem remains—how to picture the encoding of the vector potential in spacetime. Where does the vector potential live? Could it be possible to do it with three circles at each point of space-time—with axial directions in the x, y, and z directions and perhaps electromagnetic time too. If gravitational curvature is negligible, then x,y,z,t space is flat. So, not one phase along a circular fiber, but three phases with space orientations. This is also a thought from several years ago—but how to justify it? Well, if it works—$\boldsymbol{A_x \propto \partial\phi_x/\partial x}$, $A_y \propto \partial\phi_y/\partial y$, $A_z \propto \partial\phi_z/\partial z$, and ϕ potential $\propto \partial t_{EM}/\partial t$. It may be easiest to picture the axis of circular rotation to lie along the space axis direction (so A_x would use ϕ_x with axis along x but circle in the y,z plane). Maybe this could use Pauli spin matrices or quaternions? -- $n \cdot \sigma$ where $\sigma = \{\sigma_x, \sigma_y, \sigma_z\}$.

CIRCLE FUNCTIONS IN MODERN PHYSICS

DP

Abstract. A visual picture "Little Circles Model" mechanism is suggested for for electromagnetic vector potentials, $\vec{A}(x, y, z)$, and for quantum phase. This model was initially inspired by the Kaluza-Klein ('KK') proposal that there may exist a little curled-up electromagnetic fifth dimension along with the usual four-dimensional space-time. Gauge theory and Fiber- Bundle theory are similar but use circular 'internal symmetry.' KK does not address electromagnetic sources or how to go from sources to fields − for example, from current flow to magnetic fields. Also suggested here is a definition of a magnetic field, B, as a "shearing" of the A-field as if A were somewhat like an velocity field. We also discuss the $\vec{A}$ field for regions which are not simply-connected.

1. Kaluza-Klein Circles

In 1919, Theodor Kaluza submitted a five-dimensional (or '5-D') version of unified gravitation and electromagnetism to Albert Einstein for his approval. Kaluza added an extra row and column onto Einstein's 4×4 metric tensor $g_{\mu\nu}$ and inserted the EM four-vector potential A_μ [1] and an added scalar term into those locations [1]. This $A_\mu = g_{5\mu}$ row resembles the usual 4-D metric off-diagonal space-time h_{io} rows responsible for the Lense-Thirring effect also known as "gravitomagnetism" which can deflect approaching massive bodies from massive collapsed spinning stars. Kaluza restrained the influence of the fifth dimension with the assumption of a "cylinder condition." Because this 5th dimension isn't seen, Oskar Klein suggested in 1926 that it is curled up into a tiny circle so small that it cannot be directly detected [6] − a tiny cylinder condition. He gave the fifth dimension a circular topology so that its extra coordinate x_5 is periodic, $0 \leq x_5/R \leq 2\pi$ where R is the radius of the circle S^1. Thus the new space has topology $M^4 \times S^1$ alternately stated as having an S^1 bundle over 4-dimensional space-time M^4 with $x = x_1$, $y = x_2$, $z = x_3$ and time $t = x_o$ or an older view $t = x_4$. Physicist's modern 'gauge theory' and mathematician's 'fiber bundle theory' say somewhat similar things about electromagnetism. The KK cylinder condition is similar to gauge phase invariance or U(1) symmetry also called a 'structure group' for fiber bundles. In KK, the potential terms A_μ's are metric coefficients; while in gauge theory the gauge potential A is a local 'connection.' KK phase invariance is the grandfather of Yang-Mills theory and gauge theories and more recently to higher

Date: February 8, 2012.

email: davepeterson137@gmail.com. Paper updated to 1/2/14.

[1] $\mu = 0$ (for time) and 1,2,3 for x,y,z space. The older 'ict' $= icx_4$ notation has been in progressing disuse since the 1970's

dimensional supergravity and string theories. The five-dimensional Kaluza-Klein miracle was that having KK 5-D 'Ricci' flatness $R_{ab} = 0$ and circle-group 'U(1)' symmetry for x_5 gives **Maxwells equations** for electromagnetism (− but without sources of currents or charges, see 'More on KK' at end) [2]. Kaluza-Klein does not show how to go from source influence on 'little-circles' to the vector potential; the A values are simply placed into the new metric by hand. Gauge theory and Fiber Bundle theory also do not state how the little circles lead to the vector potential. That has been an unaddressed road-block to understanding. This paper discusses a visualization of this missing connection.

It has been somewhat difficult to form useful mental pictures of the Kaluza-Klein idea. In publications, only abstract mathematics is typically provided. In the simplest view, a little circle is attached to each point of space-time as in Figure 1 below (e.g., picturing an x-y lattice with circles sticking up at each intersection point). One picture for the S^1 circle in 5-D is then a vertical hoop with an advancing red phase-spot on it[3]. The plane of this hoop can be tilted so that its projection onto orthogonal planes containing axes x_5 $\&x$, x_5 $\&y$, x_5 $\&z$ can be circles with different x, y, or z component strengths or amplitudes. In the drawing in Figure 1, suppose that a current carrying wire is oriented in the x-y plane at some angle from the axes. Its current flow $I = J\mathcal{A}$ is its current density times cross sectional area often referred to as just 'source J'. From a 3-space view (E^3), one sees an oscillation or rotation of the circle along some line like this which is tilted with respect to the x, y and z axes with direction cosines, a_i (e.g., $a_1 = \cos\alpha$ between J and the x axis)[4]. This tilted line should be parallel to or follow the net effect of the currents producing the vector potential ($\vec{l} \parallel \vec{A} \parallel \vec{J}_{net}$) − 'like a flag in the wind.' Then one circle with one dot can now be seen as three circles with three red dots along the various axes. The 'tilt' idea is kind of like a 'connection' between the S^1 fiber and the x y (or z) space. This tilt essentially gives a vector field prior to differentiation. Each point of space will have a little circle like this each with a slightly different phase, φ. Figure 2 is a standard drawing (e.g., [4]) showing how the phases on these little circles or 'fibers' can change as one progresses along a path in space.

A gauge potential or local connection, A, gives a rule for lifting a curve in the charged particles position space to a curve in the U(1) bundle of position wavefunctions [8].

Such a picture enables me to think that I have some understanding of electromagnetism and perhaps some glimpse of the reality underlying other forces. In the little-circles picture, the machinery of space-time is presumed to be a complex network enabling awareness and communication between neighboring circles. Each circle is aware of and consistent with

---

[2] $R_{ab} = 0$ or $G_{ab} = 0$ in 5D implies that ${}^4G_{\mu\nu} = {}^4\ T_{\mu\nu}{}^{EM}$ in 4D, and the cylinder condition and weak-field-limit gives the classical homogeneous vacuum solution without sources J or ρ (see second half of equation (8) below). Also, the Christoffel connection $\Gamma_{5\mu,\nu} = \frac{1}{2}\kappa F_{\mu\nu}$ [12]. EM Stress-Energy $T^{(EM)}_{\mu\nu} = F_{\mu\alpha}F^{\alpha}{}_{\nu} + \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}$ (Adler I p 399).

[3] Why red? Because the laser pointers of my day produced bright red laser spots.

[4] I believe this 'Tilted Hoop' idea is my own idea bypassing 'connections' and 'fibres' and the usual overly-abstract advanced math without pictures

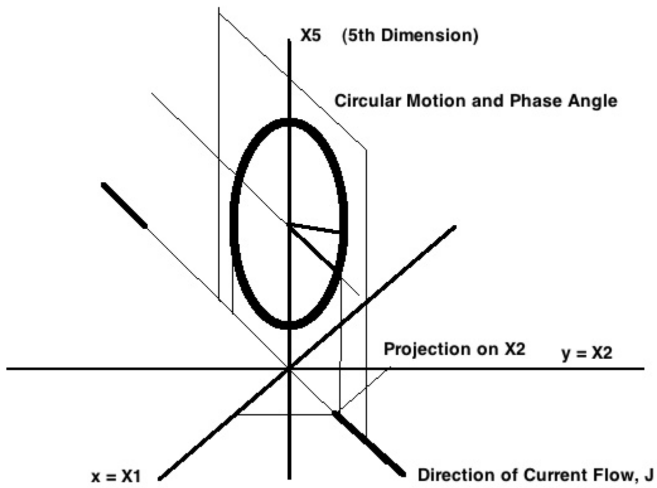


FIGURE 1. KK S^1 Circle attached to an xy point of space and oriented up in the 5th-Dimension x_5. The phase angle is φ.

its neighbors and compares their phases with its own. What we call A may be the rate of change, $A_i = a_i g \partial\varphi/\partial x_i$ [6], of 'relevant' phases of these circles with respect to their neighbors as a 'directional derivative' (not a pure gradient). The 'g' is a units-dependent relevant coupling constant for A (e.g., perhaps $\hbar/e$). The phase, φ may be like a 'sub-potential' in the computing done by Nature.

There are several parameterizations of a circle. One could refer to it by two components (x,y): $S = \{(x,y)|\ x^2+y^2=1\}, \quad x=\cos t,\ y=\sin t$. Because of the periodicity of sine and cosine, the parameter t could be chosen either as $t \in \mathbb{R}$ or as $t \in \mathbb{R}\, mod 2\pi -$ the latter choice being a compact manifold. Another choice is like the (Lie) circle group:

$$\mathbb{T} = U(1) = \{z \in \mathbb{C} : |z| = 1\}, \quad z = e^{i\theta} \tag{1}$$

My choice references a monotonically advancing phase through space or time which is more like the real line but wrapped around a circle like a compact helix. This phase will then have a "winding number" $n = (\phi_2 - \phi_1)/2\pi$. An angle phi will look the same as a 'wrapped angle' phi + $2\pi n$. In elementary math, there is a "wrapping function" $w : \mathbb{R} \to S$ such that a radian phase or arc length t on the unit circle determines an (x,y) location on the circle, $S^1 \subset \mathbb{R}^2$. $w(t) = (\cos t, \sin t), t \in [0, 2\pi); \quad w(2\pi) = (1,0),\ w(\pi/2) = (0,1)$. This is pictured as string of length t wrapping around a spool of unit radius. The action of the w function can be visualized by considering the real line as a helix of radius one and w as a projection down onto the unit circle centered on (x,y)=(0,0). Loops of the helix can be

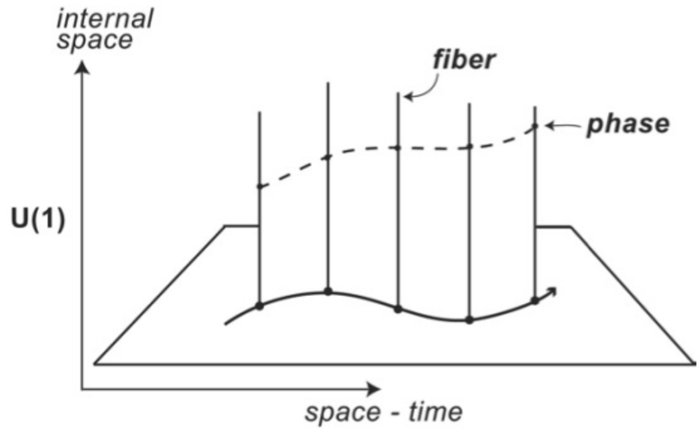


FIGURE 2. Standard picture for Phase on Fibers versus path in space-time base space $(E^3 \times R^1)-$ path dependent phase change. The fiber segments have their endpoints identified so that they represent circles.

considered one loop at a time: $\alpha(t) = (\cos nt,\, \sin nt)$ where $n \in \mathbb{Z}$ is an integer. For the whole helix, the w-mapping is not 1-1.

In a little more detail, if φ were only considered mod 2π and we look at phase differences near 2π angle, what would be the difference in phase from say $380^o - 340^o$? Obviously, we would take the difference first and then modulo the $360^o = 2\pi$. The zero point is merely a coordinate choice of no consequence. We could have looked at the problem in just x and y coordinates as a difference on the arc length of the unit circle. These are just coordinate system and zero choices of no invariant meaning. There will be few real cases where a path will be examined for more than one full circle, so the winding function concept will not often be necessary beyond considerations of multi-valued functions rather than scalar functions.

We wish to consider the 'useful' vector potentials as having the form $\vec{A} = b\hat{n}\nabla_n\varphi$, a **'directional derivative'** vector instead of a gradient of a scalar function of varying phase. A directional derivative is a gradient projected onto a certain direction by direction cosines, $\hat{n}\nabla_n\varphi \equiv \hat{n}(\hat{n}\cdot\nabla)\varphi$. The direction $\hat{n}$ unit vector is supplied by the net background or source currents $\Sigma\vec{J}$. We have a choice here between considering say three orthogonal angles versus one phase accompanied by one unit vector for direction. Because of the physics of dragging EM space in the direction of net current sources and the preference for talking about U(1) invariance, we chose one phase with one unit vector. Amplitude is then determined by phase comparisons with appropriate neighbors.

So, in summary, what I am proposing here is a "Little Circles Model" for the electromagnetic vector potential, A :

- Each point of space-time has an associated "little circle" on a 2-plane with one of its dimensions being the 5th dimension for EM. The size of the circle is unknown and will be treated as a unit circle.
- For E&M, the plane is tilted with respect to the other dimensions, x, y, z (with time, t, as a separate parameter) so that its projection on the 5x, 5y, and 5z planes yield relevant vector component amplitudes, a_i. The tilt of the plane follows the direction of current sources, $\vec{J}$.
- The unit circles each have a selected phase or dot on the circle, $\varphi = \varphi(x, y, z, t)$.
- The vector potential strength is determined by comparing phases with respect to neighboring phases, $A_i = a_i g\, \partial\varphi/\partial x_i$ (or for a single electron, $A \simeq [\hbar/e]\, \partial\varphi/\partial x$ so that there exists a 'true-gauge' wavelength of $\lambda_\varphi = h/eA$). The vector A-field is determined by the combination of the tilt of the plane (direction cosines) and the relative change of phase in base directions.
- Without sources, the natural tendency of phases is to be aligned, $\varphi(x, y, z) = \varphi_o$. Sources induce progressive dis-alignments of phase arrows.
- For QM, perhaps the little-circle is in a separate dimension 6 along with perhaps 5 (allowing common frequencies for QM$-\psi$ and massless EM fields). Then there is no projection onto space. This phase θ is then a scalar with a conventional gradient, $p \propto \nabla\theta(x, y, z, t)$.

Largely because the KK idea used an extra dimension with incredibly tiny size and predicted no new potentially measurable physics, it was gradually forgotten. But it was then strongly revived in the 1980's due to its use and generalization in supergravity and string theory. Also, with the acceptance of gauge theory and a U(1) symmetry invariance for electromagnetism, it became clear that the previously strange "cylinder condition" of Kaluza was simply phase invariance. A theory using the SU(2) Lie group would attach a sphere to each point of space-time. Now, a picture of the space time 'Vacuum' might be that of a lattice with a tiny 6-dimensional 'Calabi-Yau' ball attached to every point. Also, Gauge Theory has risen in prominence. For electromagnetism, it also uses 'little-circles', U(1), at each point of space-time. U(1) is called an **'internal symmetry space'** rather than a '5th-dimension' (but, in the absence of space-time GRT curvature, is there really a difference?). Change in phase is $d\theta = A_\mu dx^\mu$. In this paper we are primarily concerned with space rather than time.

2. The Electromagnetic Vector Potential, A:

Although electromagnetic potentials have a history of being viewed as somewhat unreal due to having gauge freedom, it is possible to imagine select choices of gauge with A following local current distributions or charges in a laboratory [non-arbitrary, 'true-gauge,' current fields, 'real potentials' [3]]. One could also include finite speed of propagation delays using Liénard-Wiechert potentials which included 'retarded time'. [e.g., [2] with SI units]:

$$A_\mu(\vec{r},t) = \frac{\mu_o}{4\pi}\int \frac{J_\mu(\vec{r'},t)}{|\vec{r}-\vec{r'}|}\delta(t' + |\vec{r}-\vec{r'}|/c - t)d^3r'dt' \tag{2}$$

As a tangible simple example, consider the vector fields $\vec{A}$ and $\vec{B}$ on the inside of a solenoid which is tall in the $\hat{z}$ direction. A solenoid is a helical coil of wire with many turns wrapped in a cylindrical column. Its internal magnetic field stores energy (of density $= B^2/2\mu$) [5]. In cylindrical coordinates (ρ, ϕ,z), suppose that only the A_ϕ component of the vector potential exists as a circular flow of A in the interior of a solenoid. The only non-zero direction cosine is $a_i \to a_\phi = a$. A_ϕ follows the flow of source current J_ϕ about the solenoid (assume a positive current flow convention). For a long ideal solenoid, the magnetic field is nearly uniform on the inside and conventionally expressed as:

$$B_z = B_o = \mu_o nI, \quad \vec{A}_{inside} = \frac{\mu_o nI\rho\hat{\phi}}{2} = \frac{B_o\rho\hat{\phi}}{2}, \quad \vec{A}_{outside} = \frac{\mu_o nIR^2\hat{\phi}}{2\rho} = \frac{B_o R^2\hat{\phi}}{2\rho} \tag{3}$$

The A field ramps up with polar radius, ρ, from a value of zero on the solenoid axis; and the interior and exterior fields match at $\rho = R$. The curl of this field is $B = B_z = (\nabla \times A)_z = 2A_\phi/\rho$ which is non-zero. [Note that although the circular A field also exists outside the solenoid, calculating its curl gives a zero B field there]. So, at this point, we have an interior curl that exists giving a magnetic field B. The mid plane of Helmholtz coils provides a similar example (and the magnetic field induction is remarkably uniform over the inner 50% of the territory).

Now consider the model suggested before describing that vector potential A_ϕ. Space-time responds to the flow of current by (somehow) advancing all the little phases in all the little circles in its interior 3-D lattice (assuming a lattice for visual simplicity). Consider a fixed time so that the phase relations are set. The A field $\vec{A} = A_\phi\hat{\phi}$ is encoded by the relative phases neighbor to neighbor about a circle. Each little phase circle sees the phases, φ, of each other circle and forms a directional derivative. One question might be, "a rotating circle has a handedness; what determines its direction of rotation?" In most cases, the magnitude of A will decrease in some direction (radial, $-\hat{\rho}$ in this case). Usually those locations closest to the current sources will be strongest. The phase advancement will determined by sources and consistency with neighbors such that its 'spin' direction 'up' is say $\hat{z} = \hat{J}_{net} \times (-\hat{\rho})$.

Advancing around a circle means advancing in phase ('ccw' positive $\hat{\phi}$, positive J_ϕ). A distance increment around a circle is $d\ell = \rho d\phi$, so let $A_\phi = ag\, d\varphi/\rho d\phi$. That is, A_ϕ can be considered as the derivative of a phase on circles about the solenoid axis.

Solving the above equation for phase gives $\varphi = (\rho A_\phi)\phi/ag$, and a plot of $\varphi(\phi)$ is a helix. Most likely, the phase on the little circles progresses around many times in traversing

[5] or with inductance L, $\mathcal{E} = L\mathcal{I}^2/2$. If driving current is shut off, this energy might be dumped as a voltage transient spike (compensating E field caused by sudden drop in B).

a full 2-pi of the solenoid phi circle. The cylindrical coordinate system is based on the polar system for which the variable $\phi \in [0, 2\pi)$ with the mod 2π understanding that $\phi = \phi_o + 2\pi n = \phi_o$. At phi = zero, the function is either discontinuous or multivalued so that the idea of the gradient of a scalar does not apply. That point is worth emphasizing. A gradient of a true scalar field, say χ rather than our φ , would have the same phi component form as the above, $A_\phi \propto \partial\chi/(\rho\partial\phi)$; but a true gradient, $\nabla\chi$ would also have a strong radial component! − which we lack here (our $A_\rho \simeq 0$). We know that the curl of a gradient of a scalar field is zero. For the solenoid case here, if the vector potential were indeed a gradient, then we would have $A_\phi = \partial\chi/(\rho\partial\phi)$, <u>and also</u> $A_\rho = \partial\chi/\partial\rho$. Then:

$$(\nabla \times A)_z = \frac{\hat{z}}{\rho}\left[\frac{\partial(\rho A_\phi)}{\partial\rho} - \frac{\partial A_\rho}{\partial\phi}\right] = \frac{\hat{z}}{\rho}\left[\frac{\partial^2\chi}{\partial\rho\partial\phi} - \frac{\partial^2\chi}{\partial\phi\partial\rho}\right] = 0. \tag{4}$$

Or, letting χ or $\varphi = B_o\rho^2\phi/2ag$ would give $(\hat{z}/\rho)[2\rho B_o/2ag - B_o\rho/ag] = 0$ again.

What happens in the real vector potential case is that **only** the first term exists and there is no A_ρ term to negate it to zero. Our real A field is NOT a gradient of a scalar field. Not only is it lacking the expected A_ρ part, but there is also no scalar field − it is multi-valued. And the vector components of 'real A' come not from gradient derivatives but rather from direction cosines of the tilt of the 2-plane containing the 'little circle.'

So, the B field in this 'real A' case comes from the "$\partial(\rho A_\phi)/\partial\rho$" portion of the curl − somewhat like the build up of A from the center of the solenoid to the coil. This bears some resemblance to the old idea of fluid shear (e.g., du/dy rate of change of wind speed near a boundary). And then one is reminded of concepts like "shear stress" $[\tau(\rho) = \mu\partial u/\partial\rho]$ with some sort of viscosity, mu; and also of an infinitesimal rotation vector or axial vector $w = \nabla \times u/2$. It could be that this outdated and largely forgotten idea goes back to Maxwell's visualizations of the vector potential, A. But, there has been a long-standing problem of finding an adequate or understandable definition of a magnetic field (beyond the phenomenology of the Lorentz force $F = qv \times B$). Here, the concept of shear from the dragging profile of electromagnetic space near flowing currents is again suggested.[6]. There is a general understanding that for fluid current flows at least, a curl exists if a little paddlewheel rotates in the current.[7] To carry this over to E&M, one must again imagine that A is similar to a flow of something. {What would be a appropriate composition for an analogous E&M paddlewheel? Oscillating charges analogous to the Foucault pedulum − but the Coriolis or Lorentz force then applies}. Also note that the absence of a B field exterior to a solenoid clearly implies that a curl is not a 'swirl' − the A-field goes

[6]Or for two parallel planes separated by some distance and moving laterally past each other, a fluid will experience laminar shear. For two current sheets flowing in opposite directions, there is also an A-shear between them. There is no swirl, but there is a magnetic field, B.

[7]There is a radially decreasing A field but no B field outside a solenoid. Would a paddlewheel rotate there?

around the solenoid but has zero curl there. So B as shear is more accurate than B as swirl.

The electric field is $\vec{E} = -\partial\vec{A}/\partial t$. This time, suppose we have a vertical ($\hat{z}$) antenna with current oscillating up and down. This current will drag a propagating A field along with it. So, let $\vec{A} = \hat{z}A_z(\rho) \simeq \hat{z}A_o \sin(k\rho - \omega t)$. The phase, $\varphi = \varphi(\rho, \phi, z)$. The electric and magnetic fields will be:

$$E = -\hat{z}\, ag\,\frac{\partial^2\varphi}{\partial t\partial z}, \quad B = -\hat{\phi}\, ag\frac{\partial A_z}{\partial\rho} = -\hat{\phi}\, ag\frac{\partial^2\varphi}{\partial\rho\partial z}. \tag{5}$$

In all cases, E and B can be generated from A which is generated from φ.

3. Quantum Mechanics and the Gradient Scalar Puzzle:

Gauge Theory makes practical use of the fact that electromagnetic (and other) potentials have freedom in choice of value and functional form − called gauge freedom. It makes impressive use of the gauge freedom of the gradient of scalar fields (the amazing relevance of irrelevance). It is as if physics forms fields from some sort of generalized curl which has no direct use for generalized gradients.

The message of EM gauge theory is that having local phase changes to a wavefunction, $\psi \to e^{i\chi(x)}\,\psi$, means that derivatives $\partial_\mu\psi$ now have a term with $\partial_\mu\chi(x)$ and that preserving symmetry [8] uses a covariant derivative $\partial_\mu \to D_\mu$ which now has to introduce a gauge field A with $\delta A = -\nabla\chi$ [11]. But this gradient is merely the un-useful part of a vector potential. The heuristics of going from this $\nabla\chi$ gradient to a fuller and more useful A is often not discussed.

The vector potential A can be altered to $A' \to A+\nabla\chi$ with no change to observable fields $B = \nabla\times A$ because $\nabla\times\nabla\chi = 0$ when χ is a scalar function [5]. In terms of 4-vectors, one can write: $A'_\mu \to A_\mu + \partial_\mu\chi$, and this preserves the anti-symmetric EM tensor $F^{\mu\nu}$. In classical physics, electromagnetic potentials are not directly seen; but in quantum mechanics they can reveal themselves through phase shifts in the wave-function of an electron. And then Aharonov-Bohm [AB] interference experiments can actually measure the relative phase shifts. In particular, for non-relativistic quantum mechanics, the Schrödinger equation including EM is shown and has a solution (supposing that potential $\phi = 0$) :

$$\frac{1}{2m}(-i\hbar\nabla - qA)^2\psi = (E - q\phi)\psi, \qquad \Rightarrow \psi(x) = \; e^{(iq/\hbar)\int A(x)\cdot d\ell}\;\psi_{(A=0)}. \tag{6}$$

In general, this is 'non-integrable' so that the phase in the exponent **depends on which path is followed**. But, if $A = \nabla\chi$, then it is trivially integrable and just gives a χ phase shift to ψ. The choice of χ could be said to specify a 'section' of a fiber bundle. The term $e^{(iq/\hbar)\int A(x)\cdot d\ell}$ is called a 'geometric phase;' and if it traverses a closed path it is called a

[8] for example for the invariance of the free particle in the Schrödinger equation

'holonomy,' $h(\gamma)$, around the path, γ. The 'action' integral $(q/\hbar)\int A(x)\cdot d\ell$ is a 'Dirac phase' [Dirac, 1931] and is closely related to the AB effect [1959]. Yang said that electromagnetism "is the gauge-invariant manifestation of the nonintegrable Dirac phase factor."

Consider again the case of the lack of a magnetic field outside a solenoid. $B = 0$ whether the current is turned on or turned off. The vector potential there is almost an example of a gradient of a scalar function, χ. It has the form: $A_{out} = B_o\hat{\phi}R^2/2\rho$. But $\nabla\chi = \partial\chi/(\rho\partial\phi)$. So $\chi = B_oR^2\phi/2$. If the current is turned on in the solenoid, then the phase of a wave function will change: $\psi = \exp(ie\chi/\hbar)\,\psi_o = \exp(ieB_oR^2\phi/2\hbar)\psi_o$. The phase changes even though any electrons present never experience any magnetic field! [10]. One does have to be careful about what happens when angle phi advances from 0 all the way around again to 2π (e.g., avoid that point for AB electron paths by starting at $\phi = \pi$ radians). This case is very interesting. The base manifold is the space external to the solenoid (or in cross-section, $M = R^2 - \{\rho \le R\}$) $-$ and of course $\chi \propto \phi$ couldn't be define at $\rho = 0$ anyway. The phase and chi are lifted above the manifold like an annular-helix with central cylindrical hole (a helical staircase or 'split-lock-washer'). It does look like a scalar field except that it isn't a function because it is multi-valued. The integral of $A \cdot d\ell$ is zero around any closed path except the path all around the solenoid itself. If electrons could go around the solenoid in different directions more than once, the split in phases would be even stronger.

Paths that do not go around the hole (the solenoid and its interior) are 'simply connected' and obey the usual Green's Theorem, $\int \nabla \times A\, d\mathcal{A} = \oint A \cdot d\ell$. In this region, $\nabla \times A = 0$, $A = \nabla\chi$, $\oint A \cdot d\ell = 0$, and $\int_p^q A \cdot d\ell$ is path independent. But, when a path goes around the hole, it is no longer true that $\nabla \times A = 0 \Rightarrow A = \nabla\chi$ as a true-gradient field. The Aharonov-Bohm Solenoid effect is due to the path not being simply connected. This is a huge difference that causes confusion in the literature and among students.

For non-relativistic quantum mechanics, suppose that QM has its own special orthogonal dimension and do something similar to the KK little circles idea. If the little-circle plane is now in dimensions 5 and 6, they could share a common frequency for the case of massless electromagnetic fields (a previous mystery). Possessing no "direction-cosines" in space, this new phase, θ will be a scalar and can have a gradient. [This phase, θ may or may not be the same as the previously used phase $\varphi-$ perhaps with a phase shift of $\pi/2$]. Wave momentum can be $p \propto \nabla\theta(x, y, z, t)$. Or, $p = \hbar k = (-i\hbar\partial/\partial x)(\psi \propto e^{i(kx-\omega t)})$ for the case of a traveling wave. Here, phase $\theta = kx - \omega t$. For ideal plane waves motion, $p = |\psi|\hbar\ \partial\theta(x)/\partial x$. The gradient is normal to the wavefronts of constant phase. Similarly, $E = h\nu = \hbar\omega = \hbar d\theta/dt$. Somehow, ψ also has to carry information about vector polarizations and spin (or spinors). Coupling the EM dimension with the QM dimension might enable some of this ability.

The Aharonov-Bohm [AB] effect for a moving electron says that beyond the phase changes associated with momentum, $\Delta\theta \simeq p\Delta x/\hbar$, there will be an additional change in phase associated with A: $\Delta\theta \simeq eA\Delta x/\hbar$ (pure mathematicians use the notation $\omega = -ieA/\hbar$. Here we assume a 'physical' simple A (sometimes called a 'current field') determined in a laboratory frame of reference and following lab currents − no gauge freedom. Then, $A \simeq ag\Delta\theta/\Delta x \simeq (\hbar/e)\Delta\theta/\Delta x$ suggests that the previously undetermined coefficient $ag = \hbar/e$. Intuitively, the AB effect is like the dragging of electromagnetic space. Can that be a real view? No. The quantum-mechanical world sees frequency and wavelength ($\omega and \vec{k}$, or energy and momentum) as primary. Velocity is not primary and is only inferred.

Apart from a simple picture of QM, none of the complexities of QM nor of its interpretations are discussed here. That is a separate and massive undertaking by the physics community with no consensus after 80 years of trying. The picture here is analog without consideration of superpositions [9]. People sometimes use the term 'quantum computer network' for the machinery of the Vacuum of space-time. But if quantum means discrete, I think the transfer of quanta only occurs at the end of a transaction. It is only "digital at the completed classical level. In-between is not digital, and the term 'quantum needs more refined definition. Psi is <u>information amplitude</u>.[10]. The quantum universe does its 'thinking at this level.

4. Discussion:

The fundamental interactions of particle physics are described by quantized gauge theories using gauge potentials. Since 1977, "many physicists have been convinced that the language of fibre bundles is the correct language to use in gauge theory. Electromagnetism (without monopoles) in this terminology is 'connection on a trivial U(1)-bundle.' A connection specifies how to transport objects such as wavefunctions, vectors, or tensors from one point of a manifold to another − from one fiber to another.

A modern abstract mathematical discussion of vector potentials may be similar to the following: Maxwell's theory of electromagnetism may be described as a circle bundle using a selected vector potential connection over space-time. His EM field is a bundle curvature $F_{\mu\nu}$ (2-form) consisting of E_i's and B_i's based on a 1-form connection $A_\mu = (\phi, \vec{A})$. In general, a bundle, $B = (E, M, \pi; F)$, is a differentiable manifold consisting of a combination of a total space, E, a base space, M, a continuous surjection mapping $\pi : E \to M$, and a fiber or fiber manifold, F or Lie group G such as G = U(1) for electromagnetism − also called a 'gauge group' or 'structure group'. A Principle Bundle has a fiber which has a symmetry group operating on it. For a point $p \in M$, a fiber over p is $\pi^{-1}(p)$. The base space is Minkowski space $M = M^4$ or $E^3 \times E^1 = E^{3+1}$ for space + time when time is a parameter. The connection field describes how neighboring fibers are related by symmetry rotations.

[9] of course a scalar phase field can have many superpositions, but Nature would have to do a lot of complex analysis to separate it into its components. Does Nature do that? Probably, and with amazingly precise efficiency, too.

[10] also misleadingly (but safely) called 'Probability Amplitude.'

Lifting a path in a fiber bundle means finding a path in the total space starting at a given point and lying directly above the corresponding path in the base space. A connection defines a path-lifting rule without needing the concept of parallel transport (as is used in general relativity). This sort of 'parallel transport' doesn't have to lift a closed curve in M to a closed curve in the bundle [8]. If the U(1) phase changes when going around a closed loop, then $F_{\mu\nu} \neq 0$. Does any of this modern differential geometry elegance provide insight into the physics of going from little-circles to the vector potential, A? No. None of the journal, texts, or web-references provide this step. The ideas of Figure 1 might aid in understanding this missing link.

In the circle manifold, one can take a derivative through the mod 2π point because a derivative only uses differences which don't depend on zero-point $-$ just on relative positions of the dot on the circle. The neighborhood differences do not care where the mod break occurs $-$ phase φ is simply always advancing; the little red dots go around the circle without caring where the 0^o origin is. If we want a variable ϕ to live on the circle, there is no discontinuity at 0 or 2π (they are identified, mod 2π). But, can I take a derivative at that point? Yes, because the absolute phase is unimportant when only differences are being considered.

A circle is one-dimensional but has a different topology from a line. It is the simplest example of a topological smooth manifold. A connected 1-dimensional manifold is either the circle (if compact) or the real line (if not). The structure of a manifold is encoded by a collection of charts that form an atlas. One example (of many) in this case is an atlas of 4 line segment charts: there are upper and lower x line segments (-1, +1) and left and right y line segment charts (-1, +1) that can be mapped from all short arcs of the circle. It is locally Euclidean and enables calculus (a differentiable structure). These charts (open intervals $\subset \mathbb{R}$) can also be conveniently pictured as the square box surrounding, touching and tangent to the unit circle. Label its sides as T (for top), R, B, and L. The top mapping is $(x, y) \rightarrow x$ and the right mapping is $(x, y) \rightarrow y$. They have an overlap of (0,1). Consider an element $a \in (0, 1)$. A transition map from T to R is from a in T to the circle and then to R: $T(a) = R(T^{-1}(a)) = \sqrt{1 - a^2}$. The circle is a differentiable manifold: it is smooth, its charts are smooth, and its mapping and transition maps are smooth.

5. Coupling Problem:

Consider an electron sitting inside a solenoid. It's effective quantum wavelength (without inertial momentum) is $p_{em} = eA = h/\lambda_p$. It should sense a spatial wavelength without moving. The phase circles also have a wavelength, $d\varphi/d\ell = d\varphi/\rho d\phi = 2\pi/\lambda_\varphi$. We have also formed an A-potential field given by (3): $A_\phi = ag\Delta\varphi/\Delta\ell = B_o\rho/2$. Therefore, the phase circles around angle ϕ have $\lambda_\varphi = 4\pi ag/B_o\rho$. If we try to equate the em-momentum wavelength with the vector potential circles wavelength, $\lambda_\varphi = \lambda_p$, then we get coupling $ag = h/2\pi e = \hbar/e$.

As an example, for a $B_o = 1000$ gauss solenoid field near a radius of 1 cm, the wavelength $\lambda_\varphi = h/eA_\phi = 2h/eB_o\rho = 2(6.6 \times 10^{-34} Js)/(1.6 \times 10^{-19} C \cdot (0.1 Wb/m^2)(0.01m) = 8.2 \times 10^{-12} m \simeq 8pm-$ well below a micron. So there are a great many waves around a circle.

Now, suppose that instead of an electron we have some other charged particle with $q = Ne$ where N is any positive or negative number. Then, $\lambda_p = h/qA = h/(Ne)A$ and the coupling now has to be $ag = \hbar/Ne$. The coupling depends on the test charge sign and strength. But the A and B field do not have this dependence. The B field doesn't depend on test charge because B^2 represents physical energy density by itself. EM momentum in the AB effect does always depend on the product of A and test charge (and for some nearly closed path). But I believe this has only been verified for electrons in real interference experiments. It may be that singling out an electron for reference is somehow valid.

KK's original intent was for the radius of the x_5 circle to represent one unit of electron charge, e. This size of charge is ubiquitous in Nature for almost all of the hadrons, leptons and mesons (an exception is the Δ^{++} hadron (uuu, 1.23 GeV) $-$ and the quark fermions with fractional charges). But I don't believe that the AB effect has even been tested for protons yet $-$ and the lab wavelength would be so small as to probably be unmeasurable. It is easy to imagine that the 5th-dimension or the inner symmetry for U(1) electromagnetism is indeed reserved by Nature just for the representation of one unit of electron charge. Electromagnetism in the classical world would not detect quantum aspects.

6. More on 'KK':

The majority of literature on Kaluza-Klein (KK) say that it unifies gravitation and electromagnetism and can output Maxwell's equations. A deeper study of the literature reveals that this isn't quite true. One clarification is that KK is consistent with the 'Einstein-Maxwell' equations. What that means is that general relativity can include an addition to the stress-energy tensor which includes the electromagnetic field 'in vacuum.' The term 'in vacuum' means that while there may be electromagnetic fields present, there aren't any source charges or currents (or magnetic monopoles).

But when we refer to Maxwell's equations, we usually mean (differential form with SI units):

$$\nabla \cdot B = 0, \quad \nabla \cdot D = \rho, \quad \nabla \times H = J + \partial D/\partial t, \quad \nabla \times E = -\partial B/\partial t \tag{7}$$

for 'no-poles,' Coulomb's Law, Ampere's Law, and Faraday's Law. Coulomb's (Gauss' law) has a charge density source, and Ampere's Law has a current density source, J. $D = \epsilon_o E + P$ and $B = \mu_o(H + M)$. In relativistic covariant form where $F_{\mu\nu} = A_\mu, \nu - A_\nu, \mu$ (with ' , ' meaning partial differentiation with respect to x $_{\text{'}GreekLetter\text{'}}$):

$$F_{\mu\nu,\nu} = \mu_o J_\mu, \quad and \quad \partial_{\,[\gamma} F_{\mu\nu]} = 0 = F_{\mu\nu,\gamma} + F_{\gamma\mu,\nu} + F_{\nu\gamma,\mu} \tag{8}$$

It is possible to derive the latter 'homogeneous' Maxwell equations (using 'Bianchi identity' in 5-space). But, the first equation comes out of KK as:

$$F_{\mu\nu,\nu} = \nabla^{\nu} F_{\mu\nu} = 0 \tag{9}$$

again showing that no sources ρ or J's are present in free space. In addition, the KK calculation gives $F^{\mu\nu}F_{\mu\nu} = 0$ implying the strange constraint that $E^2 = H^2$? This is a weakness of KK theory that remains uncorrected to this day. It was also unclear why compactified dimensions like the little circles should stay tiny. Higher dimensional KK theories had to deal with the fact that the standard model has chiral fermions (electrons and quarks have a handedness). KK was largely replaced with newer and bigger theories (1975-1978, Scherk, Schwarz, Cremmer). Attempts to patch up pure KK theories were unsuccessful until the present day application with many more space dimensions being used (11-dimensional supergravity theories in the 1980's and 10 dimensional superstrings).

Kaluza had made the assumption in his calculation of the 'cylinder condition' that the values in the fifth dimension play no role in physics and that we need not take any derivatives with respect to the fifth coordinate. Klein showed in 1926 that the cylinder condition would arise naturally if the fifth coordinate had a circular topology so that physical elds would depend on it only periodically [9]. With current understandings from gauge theory, this is no longer surprising because U(1) gauge or phase invariance means invariance with respect to coordinate transformations along the fth dimension. Kaluza inherently assumed U(1) symmetry, but Klein made it have tiny size − in part because he hoped that it might aid in understanding quantum mechanics.

As a five-dimensional theory, KK is unable to stand by itself. It has served as an intuitive basis for more advanced theories such as supergravity. But it wasn't until Edward Witten's advanced work on string theories that self-consistency was attained.
Note that the Christoffel symbol for 5-D Kaluza theory is $\Gamma_{5\mu,\nu} = (\kappa/2)F_{\mu\nu}$. [11] But the goal of the connection in curved manifolds is the parallel transport of vectors. In gauge theory, the goal of the connection potential is the translation of phase factors over paths. [10]

7. Conclusions:

This paper makes some unusual suggestions. Most authors state that the A field generates changes in phase as particles move through space. I suggest that the A field can be defined by the changes in phase over space and that particles experience these essentially pre-existing phases. Whether right or wrong, it is not clear that anyone else has provided a picture of phases and connections from sources to potentials. I am locking in the coupling

[11] In GRT, the Kaluza 5-D metric contains a row and column of A's, and the GRT 'connection' is formed from derivatives of this metric − so the Christoffel symbol uses E and B fields components rather than the different gauge or fiber connection, A.

constant to that of a single electron charge. The definition of a magnetic field given here does not seem to appear elsewhere (although I would assume that it is known to some). In spite of its importance to the Aharonov-Bohm effect, the external solenoid field is generally discussed poorly in texts and literature. The problem of 'almost gradients' for regions which are not simply connected may be known to some experts but is rare or nonexistent in literature (and sometimes even wrong). Space-time in this paper is treated as an active platform capable of processing derivatives by itself for the formation of potentials and fields. The potentials can be considered as due to laboratory charges and currents − a preferred gauge like that suggested by Liénard-Wiechert equations. While gauge theories with flexible gauges have remarkable power and guidance, they cannot claim that preferred gauges do not exist − it is an unmeasurable assumption. Potentials are now realized to have stronger reality than the previous history of 'just fields.' A preferred gauge in each special case takes that possible reality one step further. This increases the reality of phases too (but possibly beyond any actual experimental test). The claim that Kaluza-Klein gives Maxwell's equations is shown to be weak − and this is not well known.

References

[1] T. Kaluza, "Zum Unitätsproblem der Physik," Sitz. Preuss. Akad. Wiss. Phys. Math. K1 (1921) 966.
[2] John David Jackson, Classical Electrodynamics, John Wiley & Sons, Inc., N.Y., 1962 (p 464).
[3] James Mattingly, "Which gauge matters? Studies in History and Philosophy of Modern Physics 37 (2006) 243-262. Or see: http://www9.georgetown.edu/faculty/jmm67/papers/ WhichGaugeMatters.pdf
[4] K. Moriyasu, An Elementary Primer for Gauge Theory, World Scientific, Singapore, 1983.
[5] David Peterson, "Gauge Theory for Electromagnetism," November 6, 2011, unpublished PDF.
[6] Dave Peterson, "A Five Dimensional View of Electromagnetism," 9/15/05-10/17/07, unpublished PDF.
[7] Richard Healey, "Gauge Theories and Holisms , http://philsci-archive.pitt.edu/1163/1/Gauge_Theories_and_Holisms_Revised.pdf
[8] By Frank Wilczek, A. Shapere, "Geometric phases in physics," World Scientific.
[9] J.M. Overduin, P.S. Wesson, "Kaluza-Klein Gravity," Elsevier, arXiv:gr-qc/9805018 7 May1998 (102 pages). This could be the best reference article on the subject.
[10] Gregory L. Naber, "Gauge Fields in Physics and Mathematics," www.pages.drexel.edu/ ∼gln22/GaugeFields.pdf
[11] Robert Mills, "Gauge Fields," Am. J. Physics 57 (6), June 1989, pp. 493-507.
[12] Michio Kaku, Introduction to Superstrings and M-Theory, Second Edition, 1999, Springer, p. 376.

SPECIAL TOPICS

DP

1. Refraction Index for Matter Waves?

Abstract: If quantum mechanics (QM and QFT) pertains to waves and fields without objective particles, then how is it possible for an electron wave to bend to form quantized atomic-wavefunctions around an atomic nucleus? For the 'just-waves' picture, is it possible to consider the nuclear charge, Z, as conditioning the space around the nucleus so that it essentially has an 'index of refraction, n' which can bend the electron matter-wave? Due to dispersion, n depends on initial energy, $n = n(T_o, V(r))$. Also, the transition from free to bound requires a loss of energy. But with that understanding, the idea can be made to work. A localized wavefront rotating coherently about a nucleus into a standing wave requires that the energy of the wave obey the virial theorem ($KE = |PE|/2$).

The Schrödinger equation describes 'matter waves' rather than particles. In the small confines of an atom, matter wave wave-fronts must curve strongly. How can this be explained intuitively? Is it possible to discuss the matter-wave motion in a one electron atom using the analogy of an optical index of refraction, n? That was an early intention of de Broglie by analogy to light traversing a medium of variable refractive index; and there is a topic in early 'de-Broglie optics' for electron motion as matter-waves with wavelength given by $\lambda = h/\sqrt{2m(E-V)}$, where $V = e\phi$ is electrostatic potential energy. Kinetic energy can be written as $p^2/2m = (\hbar k)^2/2m = T = E_{total} - mc^2 - V$, where $k \equiv 2\pi/\lambda$. de Broglie's overall analysis came very close to declaring wave equations very similar to the Schrödinger equation. Using these equations suggests an index $n = \lambda_o/\lambda = k/k_o = \sqrt{(E-V)/E} = \sqrt{1-V/E}$. But, this is intended in reference to an initially free particle where $T_o = (KE)_o = E_o > 0$. So, this form is mainly useful for cases like electron beams in an electron microscope (but not for bound electron orbits).

As an beginning analogous case for comparison, consider the case of light rays in general relativity theory (GRT). It is possible to discuss first order gravitational lensing and the bending of starlight by replacing gravitational curvature with a flat space having an effective

email: davepeterson137@gmail.com.

Topics: Index 6/18/12- 7/24/2012 , Water Waves: 12/29/2011, 'Circles' updated to 2/15/2012. Last update was on 8/21/12, dp.

index of refraction near gravitational sources. This approach dates back to the early 20th-century and is a strong simplification of the geodesic equation for motion. The deflection of light by the sun is one of the key tests for Einstein's gravitation,
$\Delta\theta = 4M_\odot G/[c^2(R_\odot = b)] \simeq 1''.75$.
The gravitation index, n, is given for an isotropic metric by:

$$(1) \qquad n = \frac{c_o}{c'} = \frac{\sqrt{g_{\ell\ell}(\rho)}}{\sqrt{g_{oo}(\rho)}} = \frac{(1+m/2\rho)^{2+1}}{(1-m/2\rho)} \simeq (1+\frac{2m}{\rho}) = 1 + \frac{2MG}{c^2\rho} > 0,$$

where the metric $ds^2 = 0$, $c' = d\ell/dt = c/n$, $\rho = r = radius$, 'reduced mass' $m = (MG/c^2)$, and $m/r \ll 1$. As a geometrical optics analogy, a light ray offset by an impact parameter will get bent towards the gravitational source due to encountering an increasing index of refraction.

$$(2) \qquad \ddot{\vec{r}} = \nabla(-U(r)) = \nabla(c_o^2 n^2/2) = (c_0^2/2)\nabla(4m/r) = -\frac{c^2 4MG\hat{r}}{2c^2r^2} = -\frac{2MG\hat{r}}{r^2}.$$

This is similar to double Newtonian deflection where "U" is an effective potential energy. There is also a gravitational red-shift for photons moving out of a gravitational field. This means that the relation $E = h\nu$ describes a frequency versus a total energy which includes gravitational fields. One would also expect it to be able to include mass-energy, relativistic contributions, and electro-magnetic field energies too.

The example above [equation (2)] is a special case because setting the metric distance $ds = 0$ is reserved for light speed, c. In addition, the acceleration (r-double-dot) is really given by $d^2\vec{r}/dA^2$ where A is 'action' and is related to coordinate time by $dA = dt/n^2$. Since the usual planetary case of GRT has an effective index almost identical to unity, this is not often much of a correction. For a massive particle, an exact GRT equation is $d^2\vec{r}/dA^2 = \nabla(n^4V^2/2)$ [2] where V is coordinate velocity [for light $V = c/n-$ thus reducing back to equation (2)]. The potential is now velocity dependent, $U = -n^4V^2/2 = -c^2N^2/2$, and we can consider a new index, $N = n^2V/c_o$.

Atoms: As introductory background for the case of the one-electron hydrogen atom, first consider 'circular' orbitals. To see what angular momentum is doing, consider a simple case with maximum $\ell = m = n-1$, e.g., $u_{n\ell m} = u_{211} = A_2R_2(r)P_1^1(cos\theta)exp(+1i\phi)$, so that $u^*u \propto r^2exp(-r/a')\sin^2(\theta)$. The expected $L^2 = \ell(\ell+1)\hbar^2 = 2\hbar^2$, and the projection of the total angular momentum magnitude L on the z-axis in this case is $L_z = m\hbar = 1\hbar$ with $\langle r\rangle = n^2a_o'/Z = 4a_o'$. We must then be able to consider rotating plane wave is on planes that seem to pivot about the $\hat{z}$-axis but are windowed by the amplitude of $\sin\theta$ in the theta direction and $re^{-r/2a'}$ in the radial direction (with a null at $r = 0$).

'De Broglie optics' or 'matter wave optics' studies matter wave properties of propagation, reflection, diffraction, and interference. There are subfields of electron optics (microscopy, holography), neutron optics (diffraction and interferometry), and the new field of atom optics [3]. Matter wave and light wave index of refraction is defined as the ratio of the local k-vector = $k(r)$ to the free propagation k-vector , $n(r) = k(r)/k_o$. This applies to the action of a potential or to scattering by a medium.

In general, $E = h\nu = \hbar\omega$ and $p = h/\lambda = \hbar k$. Potential energy is charge times potential $V = e\phi$. The Schrödinger equation corresponding to this is , $(E - V)\psi = (p^2/2m)\psi$ but usually shown in operator form $[\hat{p} = -i\hbar\nabla]$. The non-relativistic energy is $E = \hbar\omega = V + (\hbar k)^2/2m$, so that the basic 'dispersion relation' is $\omega(k) = \hbar k^2/2m + V/\hbar$. Since this is quadratic in wavenumber k, matter wave-packets will spread. A dispersion relation is an equation describing frequency versus wavelengths for a particular medium and case.

Semi-classical physics associates 'particles' with wave-packets with both having an ordinary velocity or group velocity

$$v = v_{group} = v_g = \frac{\partial\omega}{\partial k} = \frac{\partial E}{\hbar\partial k} = \frac{p}{m} = \frac{\hbar k}{m} \tag{3}$$

This is a very general equation which also can apply for solid medium as unusual as 2D 'non-parabolic' graphene implying an 'effective mass $m = m^*$ which can very different from free space m [4]. In those cases, the 'optical effective mass' can be defined by: $m^*(E,k) = p/v_g = \hbar^2 k(\partial k/\partial E)$.

Part of 'wave-particle-duality' is the coexistence of group velocity representing particles and phase velocity representing waves. For this case of $\omega = \omega(k)$, the 'phase velocity' is then given by:

$$v_\phi \equiv \omega/k = \hbar k/2m + V/\hbar k = p/2m + V/p = v_g/2 + V/mv_g \tag{4}$$

For a 'free particle' with $V = 0$, we have the somewhat counterintuitive result that the group velocity is twice as fast as the phase velocity (so that v_ϕ can't 'catch up to' v_g).

Note that the product $v_\phi v_g = {v_g}^2/2 + V/m = (T/m)(1 + V/K)$.
This is very interesting since for atomic orbitals the potential energy $V = -2KE$, $1+V/T = 1-2 = -1$. So, the phase velocity would now be going backwards against the group velocity!

An elementary discussion of an appropriate index for the non-relativistic QM case with weak potentials and stronger kinetic energy, let $E' = T + V = E_{total} - mc^2$, and $T = p^2/2m = E' - V$. Assume base wave-number $k_o = k(V = 0) = \sqrt{2mT_o}/\hbar$ and consider E' to be roughly constant in free space and in the potential (no loss of energy so that the particle stays free). Then:

$$n \equiv \frac{k}{k_o} = \frac{\sqrt{2m(T_o - V)}}{\sqrt{2mT_o}} = \sqrt{\frac{T_o - V}{T_o}} = \sqrt{1 - \frac{V}{T_o}} \tag{5}$$

Even for deep attractive potentials, this will give a positive index of refraction, $n = n(T_o, V(r)) > 0$. The deeper the potential, the more kinetic energy will be attained.

But, the goal is to discuss n(r) for bound orbitals where the kinetic energy is related to the potential energy in a different way through the Virial theorem. To attain this, an initially free particle must lose some net energy, and that changes the problem and the equations. The typical virial case is where electrostatic potential is strongly negative and kinetic energy $T \simeq |V|/2$ so that $E' = T - V < 0$ (e.g., $-13.6\, eV$ bound electron energy). In more detail, the expected kinetic energy is $< KE >=< p^2 > /2m = +\hbar^2/2m{a_o}^2 \simeq +13.6eV$ for the hydrogen atom ground 1S state. 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:

$$\langle\psi|T(p)|\psi\rangle = (\lambda/2)\langle\psi|V(r)|\psi\rangle \tag{6}$$

where the potential V is of degree $\lambda = 1$ here.

Bound State Reference: Instead of beginning with a free particle reference, one approach is to reference from a bound state instead. Let $V_o < E_o < 0$ be a reference and consider slight deviations in radius and potential and kinetic energy from that basis. Let $\Delta V = V(r) - V_o = -b/r - (-b/r_o)$ for instance. Then

$$n' \equiv \frac{k}{k_o} = \sqrt{\frac{T}{T_o}} = \sqrt{\frac{E_o - V(r)}{E_o - V_o}} = \sqrt{\frac{T_o - \Delta V}{T_o}} = \sqrt{1 - \frac{\Delta V}{T_o}} \tag{7}$$

Then the case $r > r_o \Rightarrow b/r < b/r_o \Rightarrow -b/r > -b/r_o$ so that if r increases, then ΔV increases and index n decreases. Then consider what happens if we have a wavefront on a radial line consisting of separate little wavelets staggered at different radii each hoping to move about in a circle but all having the same energy $E = E_o$. A lower wavelet will have higher n and smaller wavelength than a middle wavelet; and a higher wavelet at higher radius will have lower n and larger wavelength. The result will be a net bending of the wavefront towards an orbit around a nucleus.

But is it the right amount of bending? To have a wavefront in circular orbit with circumference C requires that the effective $\lambda(r) \propto C = 2\pi r$. If we let $r_{upper} = r + dr$ and $r_{lower} = r - dr$, then we need $\lambda(r) = \lambda_o(1 + dr/r)$. What we have is $\Delta V \simeq bdr/{r_o}^2$ giving to first order approximation $\lambda(r) \simeq \lambda_o(1 + bdr/2T_o{r_o}^2)$. But the Virial theorem for inverse square orbits gives $b/r_o = |V_o| = 2T_o$. So the requirements do indeed match within first order approximations! Beyond that we probably lose phase. So, an effective local index of refraction can produce a rotating phase front so that a matter wave can form a circular orbit about a nucleus. The remaining mystery is what happens to the radial part of the wavefunction.

The ground state S orbital for a hydrogen atom has an exponentially decaying radial 'tent' profile: $\psi_1(r) = \frac{1}{\sqrt{\pi}}\left(\frac{Z}{a_o}\right)^{3/2} e^{-Zr/a_o}$ [13]. We have an inverse square electrostatic field which has infinite potential and force at zero radius but also a null in the 3-space probability of actually being there. The peaked tent profile is not smooth − unlike the smooth harmonic oscillator bell-shaped wave-function based on zero force at the origin.

Relativistic Case: Since it is more likely that a particle's basic vibration originates in its mass-energy, $\omega_o = m_o c^2/\hbar$, and $\omega = (Total\ E)/\hbar$ which includes mass and momentum and potentials. Then a more appropriate equation is something like the Klein-Gordon equation: $(E-V)^2\psi = [(pc)^2+(mc^2)^2]\psi-$ usually shown in operator form. The relativistic dispersion relation is:

$$\omega = \omega(k) = \frac{\sqrt{(\hbar kc)^2 + (mc^2)^2}}{\hbar} + \frac{V}{\hbar} \tag{8}$$

So, the group velocity is now $v_g \equiv \partial\omega/\partial k = c^2(k\hbar)/(E - V)$.
And the phase velocity is:

$$u = v_\phi \equiv \frac{\omega}{k} = \frac{\sqrt{(hkc)^2 + (mc^2)^2}}{\hbar k} + \frac{V}{\hbar k} = \frac{(E - V) + V}{\hbar k} = \frac{E}{p} \tag{9}$$

And this equation may be obtained more simply by: [1]

$$u^2 = (\lambda\nu)^2 = \left(\frac{\lambda}{h}\right)^2 (h\nu)^2 = \frac{E^2}{p^2} = \left(\frac{Ec}{pc}\right)^2 = \frac{E^2c^2}{(E - V)^2 - m^2c^4} \tag{10}$$

Then the product of the group and phase velocities for the relativistic case is:

$$v_g v_\phi = \left(\frac{pc^2}{E - V}\right)\left(\frac{E}{p}\right) = \frac{c^2}{1 - \frac{V}{E}} = \frac{c^2}{N^2} \tag{11}$$

where an index of refraction could be considered to be $N = N(V) = \sqrt{1 - V/E}$.
This is the same result as obtained above for the non-relativistic case, but now without having the possibility of an imaginary index, N. In most cases, the potential V is negative but much smaller than the total relativistic energy so that $N \simeq 1$. The group velocity is often low so that the phase velocity is very high and faster than the speed of light: $u = v_\phi = c^2/N^2 v_g$. Usually this is considered to be non-physical; but a standing wave would no longer be an evolving 'wave-packet.' N is an index of refraction for matter waves in the same sense that it would be for optics.
For a free particle, $v_\phi = \omega/k = (\gamma mc^2/\hbar)/(\gamma mv/\hbar) = c^2/v$.

2. Discussion and Complications

For the view that QM/QFT only deals with waves/fields and not with 'particles,' we would like to see 'matter-waves' rotating around the nucleus somehow. One picture might be restricted wave fronts of planes rotating about the proton of the hydrogen atom. That means that $v_\phi = r\dot{\theta}_o$ with some angle theta advancing at near constant wave-front rates. Then $v_\phi \propto r$ so that as r increases, phase velocity also increases. That indeed happens in both the NR-QM and relativistic case. Also, one has to be able to explain S-waves which do not orbit. Is there a turn-around when $K + V < 0$?

What is the physical meaning of the fundamental 'vibration' in the equation $E = h\nu = \hbar\omega$? As one consideration, for Bohr orbits, could it correspond to the angular speed of the rotating electron? No. It turns out that the n-th Bohr orbit will have: $E_n = \hbar(\omega_{n\,rot})n/2$. This means that the special case of n = 2 will indeed have a rotating speed about the proton nucleus that agrees with $E = \hbar\omega$. But in general it isn't true. Furthermore, the actual case using the Schrödinger equation for the hydrogen atom will have an angular momentum $L = \sqrt{\ell(\ell+1)}\hbar$ with a z-projection of $L_z = m\hbar$. So the angular frequency representing total energy is not the same as the orbital frequency. It is special. Now, we could make the case that since the phase velocity is half the group or 'particle' velocity, $2|\omega_{phase}| = \omega_{rotation}$, and then $E_n = \hbar(\omega_{n\,rot})n/2 = \hbar(\omega_{n\,phase})n$, but we are still off by that factor of n.

In more detail, for the Bohr atom, the Bohr radius is $r_n = n^2 a_o$ where the first base 'Bohr orbit' is $a_o = 4\pi\epsilon_o\hbar^2/me^2 \simeq 0.53\,\mathring{A}$. The potential energy at radius r_n is $V_n = -e^2/4\pi\epsilon_o r_n = -\hbar^2/(m{a_o}^2 n^2)$. The net energy, K + V, is half-way up at $E_n = V_n/2 \simeq -13.6\,eV$. The orbital rotation speed is found by balancing the centripetal force with the electrostatic force: $mr\omega^2 = e^2/4\pi\epsilon_o r^2 = -V_n/r_n$.

In the case of multi-electron atoms, there is a special complication of perceived nuclear charge, Z, due to **shielding**. By Gauss' Law, the inner electron shells will neutralize or shield the effective felt charge on the outer shells. Worse than that, the outer shells have inner tails that also affect the inner total charge distribution. The "Index" is not a constant field from the nucleus but is affected by all the electrons (n depends on V which weakens with shielding). What is needed is self-consistent iterations converging on a stable index profile, n(r). This is doable but tedious.
Another difficulty may be that even for a single electron atom like hydrogen, the reinforced Born rule $q(r) \propto e\psi^*\psi$ will produce what seems to be a continuous charge distribution over all radii [1]. So the outer radii may also feel a lessened effective charge from the nucleus. The formation of an index of refraction $n = n(r, \theta, \phi)$ becomes a complex matter. It is not a conditioned space from the nucleus alone. It is no longer clear that one can define an index of refraction.
Note that Feynman's book on the Path Integral doesn't often consider paths for charged particles nor discuss their interactions. How did he manage to invent QED without being able to solve the Hydrogen atom using his path integral formulation? Maybe this complexity is one reason it took so long to apply the path integral sum over histories to the solution of the hydrogen atom. In addition, it must also be true that de Broglie optics only considers experimental electro-magnetic fields. It may not apply to quantum-perturbable fields such as in atoms.

[1] Almost like the old continuous electrical fluid "fluvium" – now mentioned in the same breath along with its old friends 'caloric,' and 'phlogiston.'

3. 'Water Waves,' Dec. 2011

When searching for an easily visualized example of wave motion, one tends to think of water waves. But, as Feynman says, "they are the worst possible example, because they are in no respects like sound and light; they have all the complications that waves can have [5]." Their surface motion is neither transverse nor longitudinal but rather more like circular motion due to the incompressibility of water. Some of this complexity can be seen in physics-lab ripple tanks that can operate in several regions [7]: short wavelength capillary waves [6], intermediate capillary and gravity waves, and then longer wavelength depth dependent waves (shallow gravity waves). Capillary waves are short wavelength true ripples depending on surface tension, σ, and water density, ρ. Longer wavelengths depend more on gravity, g. In tank demonstrations, an oscillating surface paddle vibration frequency can range from perhaps 2 to 40 Hertz; and water depth, d, can be 2 to 10 mm. An ideal transition between capillary ripples and gravity waves gives a minimum phase velocity near 23 cm/s with a frequency near $f_c \simeq 14$ Hz, and wavelength $\lambda_c \simeq 1.7$ cm. These regimes have different dispersion relations, $\omega = \omega(k)$ where phase velocity (or 'celerity, c') $c \equiv v_\phi = \omega/k = f\lambda$, and group velocity $v_g = \partial\omega/\partial k$. In the shallow gravity dominated waves, the dispersion relation was found by Lamb in 1932 to be:

$$\omega^2 = g\,k\,\tanh(kd), \tag{12}$$

where d is the depth of the water. For long wavelengths ($\lambda > 11d$), $\tanh(kd) \simeq kd$, so the shallow water approximation is:

$$\omega^2 = k^2\,gd \qquad or \qquad v_\phi = v_g = \omega/k = \sqrt{gd} \tag{13}$$

Tsunamis (seismic sea waves) have very long wavelengths which are much greater than typical ocean depth so that they are really shallow ocean waves obeying this equation [8]. For example:

$$c = \sqrt{gd} = \sqrt{9.8m/s^2 \times 4600meters} = 212m/s = 760km/hr = 472mph. \tag{14}$$

The period of these waves might be 20 minutes with wavelength near 200 kilometers. So these simple water tank equations can apply also to large scale waves.
In the deep-water approximation, kd is high and $\tanh(kd) = 1$ so that $\omega^2 = gk$.
The wave speed is $v_\phi = \omega/k = \sqrt{g\lambda/2\pi}$ with $v_g = v_\phi/2$. This means that longer waves travel faster. As an application for ocean waves with wavelengths 233 meters, $c = \sqrt{(9.8)(233)/2\pi} = 19.4m/s$.

As surface tension becomes more significant and depth is relatively very high, "gravity-capillary" waves have a dispersion relation like:

$$\omega^2 = |k|\,\left(g + \frac{\sigma k^2}{\rho}\right). \tag{15}$$

Gravity waves are again a good approximation when wavelength is large, $\omega^2 = kg$. But for short waves (say 2 mm or so), the second term dominates so that $\omega^2 \simeq \sigma k^3/\rho$. The group velocity is $v_g = \partial\omega/\partial k = 1.5\sqrt{\sigma k/\rho} = 1.5 v_\phi$. So, unlike the gravity waves where $v_g = v_\phi/2$, in this case an individual front wave will grow moving into the group and then disappear at the back of the group. A special case of interest is when the phase velocity $v_\phi = \omega/k$ has equal contributions from gravity and surface tension. This also means that the phase velocity is at a minimum with a critical wavelength given by $\lambda_c = 2\pi\sqrt{\sigma/\rho g}$. This is the formula that gives the 1.7 cm wavelength. The surface tension of water varies strongly with temperature from about 74 dynes/cm near 10 degrees Celsius to about 60 near boiling. Or, in SI units, water surface tension $\sigma(30^o) = 0.0712\, N/m$ with a slope near $d\sigma/dT \simeq 2\times 10^{-4}\ N/m^oC$.

Can there be a classical analogy to gauge theory? Consider the capillary wave case above where ripple tank wavelengths are very short (e.g., $\lambda_o \simeq 0.7cm$). Phase velocity can be altered locally from some mid range temperature by quickly heating or cooling local regions of the water. The ripple tank frequency can be fixed [e.g., $\omega = \omega_o \simeq 227 rad/s \simeq 2\pi(36Hz)$] and water density is fairly constant with temperature ($\rho \simeq 1000\, kg/m^3$ within about 1 percent). So $\sigma k^3 = \rho\omega_o^2 \simeq cnst.$, and $dk/d\sigma = -k/3\sigma$. At a given time (e.g., $t = 0$) and position, the phase of the wave is just $\varphi = kx = 2\pi x/\lambda$. The effect of a local spot of heat is to lower the σ, decrease the speed, increase the wavenumber k, and decrease wavelength $\lambda = 2\pi/k$. But, this phase change is not due to the gradient of any scalar field.

For the gauge theory of electromagnetic potentials, the A-field can add to effective momentum ($p' \rightarrow p - eA$) and hence change the effective value of k which then alters the phase of the wavefunction. In the ripple tank, k is altered by effective scalar fields of $\sigma\ or\ T^o$. There is no vector field here, for that we need a rotating fluid flow (like the Aharonov-Bohm A field around a solenoid − and this has actually been done by Berry [12] using water rotation about a drain hole). However, a ripple-tank double slit experiment can have phase shifts due to differential heating of an upper path versus a lower path. And a net phase difference over a closed path can result differential heating of closed reflected paths over time of travel. But gauge theory uses transformations of fields using gauge functions ($A' \rightarrow A + \nabla\chi(x,t)$). In 4-dimensional space-time, it is very hard to form a scalar gauge function, χ , such that its gradient resembles anything like a fluid flow. The mathematics of gauge theory treats chi as if it were a rotating phase, $\varphi = e\chi/\hbar$ 'modulo' (2π) radians. Nothing like this can be done in a ripple-tank. It seems to require internal symmetry spaces (perhaps quantum mechanics with higher dimensions).

References

[1] Chris Clark, "A theoretical introduction to wave mechanics," 17 Sept. 2010, www.dfcd.net/ articles/wavemechanics.pdf

[2] James Evans, et. al. (Alsing, Giorgetti, Nandi), "Matter waves in a gravitational field: An index of refraction for massive particles in general relativity," Am. J. Phys. 69 (10), October 2001, p. 1103.

[3] Martin Wilkens, "De Broglie Optics, Atomic, Molecular, and Optical Physics Handbook, ed. G.W.F. Drake, AIP Press, N.Y., 1996.
[4] Victor Ariel, Amir Natan, "Electron Effective Mass in Graphene," arXiv:1206.6100v1 [physics.gen-ph] 26 Jun 2012.
[5] Feynman, Leighton, Sands, The Feynman Lectures on Physics, Volume 1, Chap. 51.
[6] Wikipedia.org Capillary_wave,
[7] Wave Phenomenon: Ripple Tank Experiments, http://www.physics.utoronto.ca/ phy225h/experiments/ripple/RIPPLE.pdf
[8] Tom Garrison, Oceanography, An Invitation to Marine Science, Sixth Edition, Thomson Brooks/Cole, 2007.
[9] T. Kaluza, Zum Unit atsproblem der Physik, Sitz. Preuss. Akad. Wiss. Phys. Math. K1 (1921) 966.
[10] David Peterson, "Gauge Theory for Electromagnetism," November 6, 2011, unpublished PDF.
[11] Dave Peterson, "A Five Dimensional View of Electromagnetism," 9/15/05-10/17/07, unpublished PDF.
[12] Berry, M V, Chambers, R G, Large, M D, Upstill, C, & Walmsley, J C, 1980, Eur.J.Phys 1, 154-162, 'Wavefront dislocations in the Aharonov-Bohm effect and its water wave analogue.' Accessible with pictures at: http://www.phy.bris.ac.uk/people/berry_mv/the_papers/Berry096.pdf . A strobe set to the wave frequency was used to freeze wave motion.
[13] Dave Peterson, "Explaining S Orbitals," unpublished LaTeX PDF, August, 2012.

Other Considerations: The index $n = i = \sqrt{-1}$, and this would pertain approximately to all of the hydrogen "orbitals" (quantum number, n) . Although rarely encountered in classical optics, a purely imaginary index of refraction may not be necessarily bad. This is the common case where there is total internal reflection and evanescent wave with exponentially decaying amplitude $-$ as in tunneling. That is very much what we have for 1s-waves, purely exponential decay with radius all the way from the central nucleus. Of course, n imaginary means that k is also imaginary so that $e^{ikx} \rightarrow e^{-kx}$ for exponential decaying amplitude with distance.

AN INTERESTING FUNCTION

Dave Peterson

September 30, 1983

The function considered in this note is remarkable in that: it is very simply expressed; it possesses ornate contours; it has unusual discontinuities; and it has physical significance.

If electric charges are located at $(x,y) = (\pm a,0)$, the interaction portion of the energy density of their electric fields is given by,

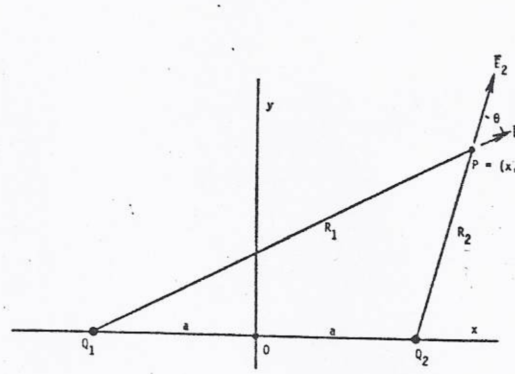
$$\rho = \epsilon_0 \vec{E}_1 \cdot \vec{E}_2 = \frac{Q_1 Q_2 \cos\theta}{16\pi^2 \epsilon_0 R_1^2 R_2^2} \qquad (1)$$

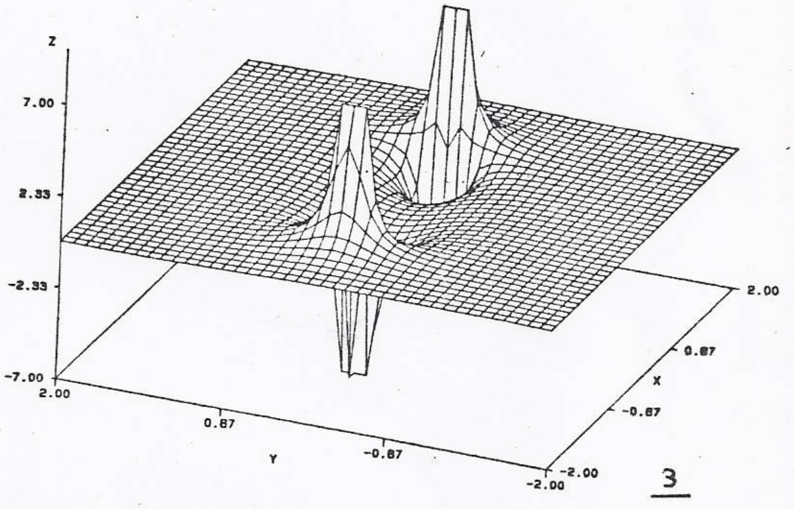
where $R_1 = \sqrt{(x+a)^2 + y^2}$, $R_2 = \sqrt{(x-a)^2 + y^2}$, and θ is the angle between the electric field vectors at (x,y) -- see Figure 1. An energy density equation of this form would also apply to the Newtonian gravitational field of two point masses and the magnetic fields of two magnetic poles. Here, we will consider the simpler function common to all three physical cases,

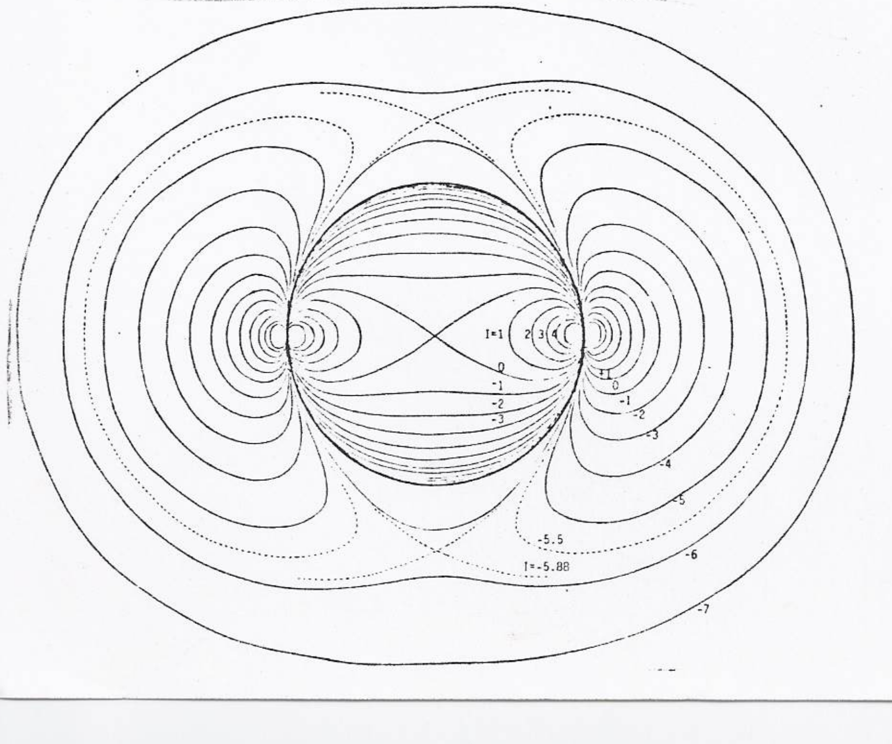
$$z = \frac{\cos\theta}{R_1^2 R_2^2} \quad \text{or} \quad z(x,y) = \frac{r^2 - a^2}{\left[(r^2+a^2) - 4a^2x^2\right]^{3/2}} \qquad (2)$$

where $r^2 = x^2 + y^2$. The contours of this mapping are shown in figure 2. We have arbitrarily selected contours at $z = \pm 1.75^I$ where $I = \{-7,-6,\ldots +4,+5\}$, and $a = 1$. Two additional dashed curves are shown for interest at $I = -5.5$ and -5.88. The circle centered at the origin and passing through the point sources at $(\pm a,0)$ represents the boundary between regions of positive and negative energy density. Near either point of discontinuity, the limit of the function is plus infinity if approached from outside of the circle, minus infinity if approached from inside the circle, and zero if approached along the circle. A three dimensional mesh plot showing the details of the topography at these points is given in figure 3 [1].

FIGURE ONE







FIGURES AND REFERENCES

FIGURE 1. Coordinate system and variables for vector fields from two point sources. The energy density at P is proportional to $\vec{E}_1^{\,2} + \vec{E}_2^{\,2} + 2\vec{E}_1 \cdot \vec{E}_2$.

FIGURE 2. Selected contours of the interaction portion of the field energy density from two point sources. Curves are labeled by values of the exponent for $z = \pm 1.75^I$. The region interior to the circle has negative density.

FIGURE 3. Three dimensional computer plot of interaction energy density arbitrarily truncated at the values $z = \pm 7$. Compare with the topography of Figure 2.

REFERENCE 1. From D. Grass, Louisville, Colorado, using "SAS" Software and an HP 9872C Plotter.

Simple Mnemonic Device for Nuclear Shell Filling

David L. Peterson

University of Colorado, Boulder, Colorado

to Am. J. Phys

Many nuclear physics texts give an energy ordering of nuclear states according to the well known shell model,[1,2] but the ordering differs in most of these texts due to dependence on A, Z, and other parameters. There is a general ordering scheme relatively common to all of these sources, however; and it can be easily pictured and remembered by students by means of the simple mnemonic device shown below.

The heavy black lines ▌ indicate the location of large energy gaps and correspond to the magic numbers 2, 8, 20, 28, 50, 82, and 126. These locations do not occur at the same place--such as after the p's in the atomic case--but they nearly do. Another difference from the electron filling scheme of atoms is the location of ▌'s in the middle of a symbol. This is due to energy depending much more on l and s coupling than in the atomic case. The l subshells are broken down into two parts by $j=l\pm 1/2$. The higher j value corresponds to the lower energy and the lower j to the higher energy--as usual the degeneracy is $2j+1$. This scheme has been given to physics students with success.

Figure 1 is equivalent to the following

Shell-filling scheme : Fig. 2. $(n\, l_j)$

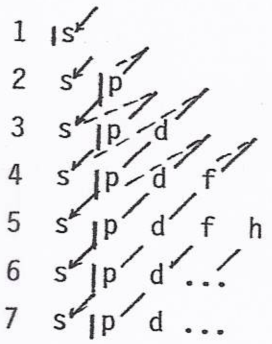
Energy ↓ $(n\, l_j)$	(number in each level)	
$1s_{1/2}$	2	2 (Magic number)
$1p_{3/2}$	4	
$1p_{1/2}$	2	8
$1d_{5/2}$	6	
$1d_{3/2}$	4	
$2s_{1/2}$	2	20
$1f_{7/2}$	8	28
$1f_{5/2}$	6	
$2p_{3/2}$	4	
$2p_{1/2}$	2	
$1g_{9/2}$	10	50
$1g_{7/2}$	8	
$2d_{5/2}$	6	
$2d_{3/2}$	4	
$3s_{1/2}$	2	
$1h_{11/2}$	12	82
$1h_{9/2}$	10	
$2f_{7/2}$	8	
$2f_{5/2}$	6	
$3p_{3/2}$	4	
$3p_{1/2}$	2	
$1i_{13/2}$	14	126
$1i_{11/2}$	12	
$2g_{9/2}$	10	
$2g_{7/2}$	8	
$3d_{5/2}$	6	
$3d_{3/2}$	4	
$4s_{1/2}$	2	
$1j_{15/2}$	16	184

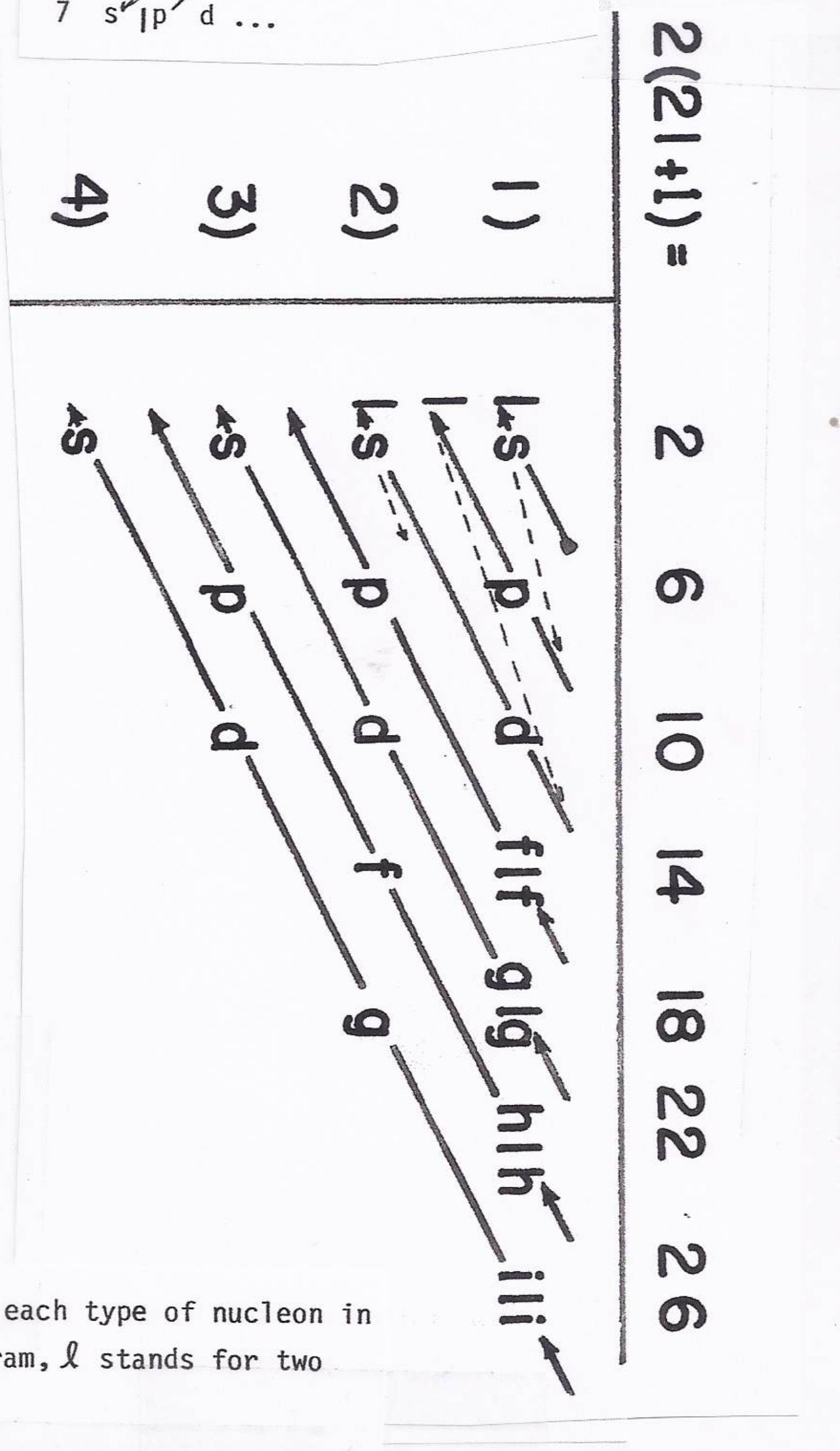
n = the no. of nodes in the nuclear wave function.

Ref. 1. Leighton - Pr. Mod. Phys

Ref 2. Evans - The Atomic Nucleus - pg 362.

ELECTRON SHELL FILLING SCHEME FROM CHEMISTRY:





2(2l+1) gives the number of each type of nucleon in each l state. In the diagram, l stands for two states: $l = (l_{l-\frac{1}{2}}, l_{l+\frac{1}{2}})$.

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]:

$$\rho = \begin{pmatrix} a_{11} & a_{12} + ib_{12} \\ a_{12} - ib_{12} & a_{22} \end{pmatrix} \to \begin{pmatrix} A & B + iC \\ B - iC & 1 - A \end{pmatrix}, \ e.g., \ \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{1}$$

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, $0 \le \lambda_k \le 0$ or $\rho \le 1$.
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 $\lambda \ge 0$ requires that: $(A - \frac{1}{2})^2 + B^2 + C^2 \le \frac{1}{4}$ [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.

[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.

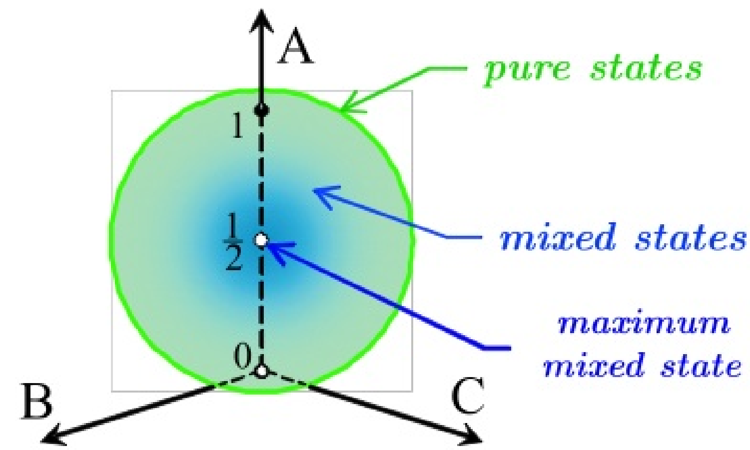


Fig. 1. Geometrical interpretation of the 2×2 density matrix.

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,

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad \rightarrow \quad \rho_{mn} = \langle u_m|\rho|u_n\rangle = \sum_i p_i \langle u_m|\psi_i\rangle\langle\psi_i|u_n\rangle. \tag{2}$$

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, $\rho_A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$.

A and B are highly correlated, but A and B by themselves are random.

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.