Boundaries

Proceedings of ANPA 24

K. G. Bowden, *Editor*

The Alternative Natural Philosophy Association

Boundaries: Scientific Aspects of ANPA 24/Keith G. Bowden, *Editor*

The Proceedings no longer has an ISBN

published by ANPA c/o Dr. Keith G. Bowden,
139 Sandringham Rd, Barking,
Essex IG11 9AH, UK

Scientific Aspects

Contents

Contents

Editorial ver 2.3

Keith Bowden,
Theoretical Physics Research Unit, Birkbeck College,
University of London, Malet St, London WC1E 7HX

Whoops, There Goes Gravity (with apologies to Eminem)

I have not spoken at ANPA for some time, although I intend to do so this year. It might be appropriate if I said something about my thinking in advance of the meeting. For about ten years now my main focus has been on the formulation of a generalised, background-independent holographic principle. (I deliberately use lower case.) This idea was put into my head in 1992 by Maurice Jessel, and clarified more recently by Pierre Noyes and Stephen Wood who suggested that it might have something to do with quantum gravity. It was also one reason why I moved my base to Birkbeck College as the idea seemed to have some connection with the work of David Bohm and Basil Hiley.

What do I mean by a **generalised** holographic principle? Well my original idea was to find a formulation of holography that was general enough to include both the Jessel Principle of Secondary Sources and Kron's Method of Tearing - which I saw as an essentially holographic process - and of course the original work of Gabor. More recently I have attempted to incorporate the link and statistical theory of Tom Etter and have thought somewhat about the work of Bohm on the implicate order and t'Hooft and others on quantum gravity. It would be particularly interesting to find a formulation that would apply to Rafael Sorkin's ideas as described in a recent meeting of ANPA.

What do I mean by **background independent**? Well the relationship between Kron's Method and Jessel's Principle illustrates this quite well. Engineers refer to this relationship as the Finite Difference, or Finite Element Method. In Kron's terms the solution to a field problem, eg an electrical potential problem described by the Laplacian (find the voltage over a continuous region whose boundary conditions are given), can be modelled - to an arbitrary degree of accuracy - by replacing the field over the region by voltages on an electrical network whose boundary conditions are given by those of the original problem.

In some way the **small scale topology** of the problem has been factored out and the solution to the problem depends only on the global topology (see below), the boundary conditions and the form of the "connection", ie the Laplacian in this case. Kron built similar network models for Maxwell's equations, Schrodinger's equation and many engineering problems. More recently I have become aware how – astonishingly – problems in fields as diverse as Knot Theory and Economics fit well into the same formulation.

On the other hand the **large scale, or global** topology of the problem must also be factored out. The idea is to obtain a holographic principle independent of all these things. Electrical circuit analogies such as the above are usually based on planar, or flat circuits, those that can be embedded in the plane. In general electrical circuits have crossovers, points where one wire crosses another. These circuits can only be embedded in higher genus surfaces such as a Mobius strip, a Klein bottle, or the surface of a torus (which may have many holes). (These embeddings are called triangulations.) Corresponding continuous field problems may be solved on the same surfaces. Similarly Physics problems in Relativity and Quantum Mechanics can only be described on surfaces that are not flat.

Think about field contours on a plane surface. The only condition on these is that they must not cross. Now think about such lines on the surface of a torus unwrapped into a rectangle in the usual way (the top and bottom of the rectangle are considered to be joined as are the left and right hand sides). Now the lines are constrained not only to cross each other but must also match at the top and bottom of the rectangle and at the left and right hand sides. Similar conditions would hold for the Mobius strip except that now the conditions at the top and bottom are relaxed and the conditions on the left and right hand sides are reversed. It turns out in general (and in higher dimensional surfaces in still higher dimensional problems) that the conditions on these contours are (perhaps surprisingly) encapsulated by a collection of mathematical (homology) groups, H_i associated only with the global interconnectivity of the surface, twists and holes. It should be possible to factor these out of any generalised, background-independent holographic principle.

I imagine the formulation of this generalised principle to live in the world John Baez refers to as higher algebra, that of category theory, sheaf theory, homology and topoi. Category theory in particular, which is the language of sheaves, homology and topoi, is a mathematical language that is generalised enough that it does not have to specify which primitive

mathematical structures, sets, groups, rings, vector spaces etc, we are working with. Kron, Jessel and Gabor etc, must all be special cases of the formulation. Kron's work had already be set in this context by the mathematicians J P Roth and Sun Ichi Amari amongst others.

Recently I put in a lot of work to understand more clearly the way the mathematics, particularly that of sheaf theory, is commonly applied to Physics problems. I was interested in the relationship between the way that sheaf theory describes structures by coverings with intersections and the intersections between torn structures in Kron's method. I am particularly grateful to Ioannis Raptis of Imperial College for pointing the way. It was not until I looked again at John Baez's excellent book in the Knots and Everything series, however, that I properly understood how such problems are described by connections on sections of sheaves and how the conditions on those connections are related to holography. I can't recommend this book enough; it is the *only* one that I have found that explains connections clearly and in a generalised abstract framework.

However I now found myself somewhat stuck. I felt that I understood most of the relevant mathematics quite well, except for topoi, (and I did not want to invest still more time into this until I felt fairly certain that I was going to get something out of it). But all my attempts to generalise the equations that I already had for Kron and Jessel were still resulting in only partial success. I wanted a new way forward. I was inspired to go back to thinking about physical holography rather than its abstract mathematical counterpart. I also started to go back to thinking about the way in which information travels around in physical problems such as those that we have been describing. It too is subject to the homological restrictions induced by the twists and holes of higher genus surfaces.

Physical holography works by phase cancellation as Peter Marcer reminds us regularly. However, I realised, I have had, in the past, a tendency to ignore the wavey properties of physics and simply think of fields. Technically, I was interested in the topological structure of problems, but tended to bury the time variation out of sight by making the homological coefficients into (modules of) polynomials in the Laplace transform. In electrical network terms, I had treated the impedances as if they were pure resistances; that is I had ignored the reactances. Perhaps this was inappropriate.

On the other hand this phase cancellation can be thought of as a restriction on the way in which information can travel round on the

surfaces. These restrictions are described technically as transmission zeros. It turns out that mathematically these restrictions have a homological structure in themselves. This was the proposal of my very first (non-computing) published paper in 1980 and one of the main themes of my PhD thesis in 1983; I had not thought about it very much in nearly as long! I will not talk about this at length this year but it is a subject that I will come back to in the future.

These transmission zeros can be much more clearly described if we work in electrical network theory rather than continuous field theory, remembering all the time the way in which the famous de Rham isomorphism (between singular homology - networks - and that of differential forms - fields) holds these two things tightly together. I was starkly reminded at this stage of a fascinating (almost) isomorphism that Lou Kauffman had shown me between electrical networks and Knot Theory. It is described in an Appendix to his book Knots and Physics, where it is somewhat hidden away. It is instructive to rewrite this in terms of Kron theory when slightly different versions of his knot invariants appear.

In this isomorphism a planar electrical network with only unit positive and negative resistances is replaced by the projection of a knot into the plane with (positive and negative) crossovers by replacing each wire by a pair of strings and each positive (negative) unit resistance by a positive (negative) crossing of the strings. Technically we get a checkerboarded knot whose dual is isomorphic to the dual of the network. If the conductance between any two points of the network is zero then the knot will pull apart if grasped by the corresponding two points and vice versa. This idea is a consequence of the knot invariants mentioned above in which conductance is shown to be an invariant of the "electrified" knot. The Reideister moves of knot theory correspond to the Thevenin and Norton star delta transformations of electrical network theory.

Recently I pointed out to Lou that there may be some physical understanding to be gained by replacing his positive and negative resistances by normalised inductances and capacitances. Inductances correspond to positive crossings and capacitances to negative crossings in the usual way. If the conductance of an electrical circuit is zero then no information can be transmitted across it. If the corresponding knot pulls apart in your hands then no information can be transmitted via the knot.

At some frequency w we can write inductance and capacitance as jwL and 1/jwC in the usual way with the electrical engineer's j being the square root of –1. Normalising and being lazy with dimensions we put w=L=C=1 giving unit inductance j and unit capacitance –j. Being slightly less restrictive we see that it is only necessary to put $w^2=1/LC$. In general we like to write the inductance as s and the capacitance as –s, considering s as the Laplace transform. All impedances in the network then take these values. Note however that it is not necessary that $s^2=-1$.

Now we can see clearly on the one hand how phase shifts can be created by the reactive components and result in phase cancellation and transmission zeros necessary for holography, and on the other how the twists in the corresponding strings may or may not result in phase cancellation and knot separation. Note however that there is a kind of double cover between the two phase shifts. A 180 degree shift in the string phase corresponds to a 90 degree shift in the electrical phase.

Remember also that this knot holography may be taking place on the surface of a torus or Klein bottle. Lou calls such knots, virtual knots. However now we notice another subtle difference between the knot theory and the circuit theory. Knots on a torus may have a single strand passing through the hole. Knots of circuits must have an even number of strands through the hole by definition. There is a similar lack of iso between the Reidemeister moves and the Norton and Thevenin star and delta transformations that wants investigating further. These problems can be easily got around, but they are also of interest in themselves.

So finally we are back at background independence and generalisation. A formulation that would include knots would be interesting and helpful. Indeed there is a natural way of doing knot homology that arises from the above; I do not know if this has been treated in the literature. This project is obviously still underway. For me the attainment of the goal is not necessarily the main point, it is the journey that has been both enjoyable and instructive. Indeed I think it quite likely that someone else will point out to me how the thing should be done (if they haven't already done it themself!) From ANPA's point of view it would be interesting to investigate to what extent these ideas are related to the Combinatorial Hierarchy. I believe that the work of John Amson, in particular, may provide some clues, but they are the subject of another discussion. I hope to explain more clearly how things are progressing in the meeting. Holograms occur on boundaries. I am using this article as an excuse for the title of this volume. Now you know why it has no dust cover.

References

Bowden, Keith (1980). *An Introduction to Homological Systems Theory: Topological Analysis of Invariant System Zeros.* Matrix and Tensor Quarterly, Vol 31, No 2, pp 37-52.

Bowden, Keith (1983). *Homological Structure of Optimal Systems.* PhD thesis, University of Sheffield, Dept of Control Systems Engineering.

Bowden, Keith (1990). *On General Physical Systems Theories.* International Journal of General Systems, Vol 18, No 1, pp 61-79.

Bowden, Keith (1993). The Spatial Transmission of Information. Proc. ANPA 15, Cambs.

Baez, John and Muniain (1994). *Gauge Fields, Knots and Gravity.* World Scientific.

Jessel, Maurice (1962). *Contribution aux Theories de Principe de Huygens et de la Refraction.* Theses pour DSc, Faculte des Sciences de la Universite de Paris. (Supervised by Louis de Broglie.)

Jessel, Maurice and Resconi (1986). *A General System Logical Theory.* International Journal of General Systems, Vol 12, pp159-182.

Kauffman, Lou (1993). *Knots and Physics.* World Scientific.

Kron, Gabriel (1963). *Diakoptics: The Piecewise Solution of Large Scale Systems.* MacDonald.

Raptis, Ioannis (2003). *Private communication.*

Smolin, Lee (2000). *Three Roads to Quantum Gravity.* Weidenfeld and Nicolson.

ANPA Proceedings Editorial Policy

ANPA has been criticised in the past - in particular by members of its own Advisory Board - for having no formal editorial policy for its Proceedings. This has been balanced by a feeling within ANPA that we should keep ourselves open to all viewpoints. In the last few years as editor I have tried to tighten things up in such a way as I felt would satisfy our critics whilst not compromising our own position. This has been partially successful although for some time I have felt that it is time that there was a formally stated policy. The following has been approved by the Executive Council, although it is open to feedback from all. By "the editor" is meant the Editor or (an) appropriate nominated Referee(s) (note the capital R!)

1. The paper should make a new and original contribution to the fields of ANPA's interest. Survey papers are acceptable.

2. The default use of language for submitted papers in Physics {and Philosophy of Physics}* should be the common language of Physics as usually understood by Physicists {and, in particular, by Philosophers of Physics}*. Any other use of language should be carefully explained at the start of the paper and all appropriate definitions included there.
{* added by KGB}

3. The editor should be satisfied that the paper is ***presented*** in such a way that the majority of the readership will understand the author's intentions. In particular ***it should be clear*** that the author has a correct understanding of the subject matter.

4. "Verbatim" reports will be accepted subject to the above three conditions only, regardless of whether the final draft is an accurate rendition of what was originally said. Other such reports are better submitted to the Newsletter.

5. Theories of any nature are acceptable material, provided they are compatible with the known facts, and provided they are deemed to be of interest to the readership. Theories of alternative, imaginary worlds are also acceptable, provided their nature is made clear.

ANPA Proceedings Notes for Authors

I would like to try to continue conformity of *style* for future issues of the Proceedings. Ideally I would like contributions to be submitted in International Journal of General Systems format (I have some copies of their Notes for Authors) or similar - **LOOK AT MY PAPER IN A RECENT ISSUE OF THE PROCEEDINGS FOR AN EXAMPLE.**

At least, Times Roman, 12 point, ***single sided, two copies (HARD COPY),*** is preferred. **10 point is TOO SMALL to be reduced to A5; 14 point is better for most papers.** Main heading 20 point capitalised and centred, other headings 16 point capitalised to the left. Author's name(s) capitalised and centred. Address italicised and centred. No underlining. At least a one inch bottom margin for footers; page numbers NOT top centre. *Only copy in good English will be considered, and remember, this is a formal Proceedings.* **Remember also to include your name (surprising how many people omit this!), affiliation and full address, email address and the version number (even if it is 1.0) or date of the draft, centred below the main heading.** I often get sent more than one version of a paper and invariably mix them up! Send copy **to *KEITH BOWDEN, 139 SANDRINGHAM RD, BARKING, ESSEX IG11 9AH.***

If English is not your first language try to keep the paper short.

The copy date for the ANPA2002 Proceedings is January 1st 2003. The issue will go to print on April 1st 2003. This will be adhered to rigidly this year.

Keith Bowden,
Theoretical Physics Research Unit,
Birkbeck College,
Malet St,
London WC1E 7HX, UK.

k.bowden@physics.bbk.ac.uk

Tel: 0208-594-5064

FOR EDITORIAL ADDRESS SEE THE PENULTIMATE PARAGRAPH

1

Some ANPA Pre-Prehistory

Ted Bastin

To Keith,

Clive and I thought this addendum to your pre history would be a good thing. I do not know what you think nor whether if you approve, in what form it should go in the Proceedings.

From Ted.

Dear Ted,

Thank you for your kind comments (below). I am sure that any sensitivity shown is more likely due to the fact that the paper was cobbled together from interviews with Clive and Pierre amongst others, then any such quality on my part! Your update is much appreciated. I think the best way to present it is verbatim so here it is.

Best Wishes,

Keith

Letter about history to Keith

Dear Keith,

I have been rereading your "Some ANPA prehistory..." in "Implications" ANPA 22. First may I thank you for being so sensitive and accurate about the personal relationships back then. Having said that I think that a bit of pre-prehistory might help people understand the picture better. You jump about ten years from the salad days of me and Clive to the catastrophic impact of Frederick.

You describe the time when Frederick didn't come to work on the PACE analogue computer at Brussels because of flu. Pace was meant to be a more reliable form of the Pask Machine. The story of this wondrous machine is a yarn well worth telling some time, but for now I will restrict myself to its logic. Gordon (Pask) provided units each of which was a massive piece of 1940's (valves) servo machinery out of a Lancaster or something to represent the first two levels of a hierarchy. Its units were in two levels with 4 and 16 of these elements respectively. It was essentially discrete since the servos worked on/off switches. The working of them was not digital but used feedback loops which connected the 16 units with each other via the 4 units. The connexions were on/off however. Gordon thought that if you connected everything with everything else while observing their total symmetry you were representing everything that could happen so that nothing arbitrary was introduced by departing from digitalness. This principle appeared in all Gordon's conceptual or actual learning and teaching machines. Input to the 16 level was randomised because though we thought the levels would continue upwards, even

Gordon could not contemplate 256 units. (4+16 units took about 2kW.)

You see that two principles - that of levels, and that of each new level's being constructed out of the operations on the existing level - already gave a hierarchy picture which Frederick used as his terms of reference. What he did was to elaborate the notion of operator in the way we all know, and if that, and the introduction of a separate mapping space, was not a stroke of genius then I do not know the meaning of that word. The suggestion which has gone about, that I asked him to produce some numbers, which he did by fiddling about with bits of algebra, does not do justice, particularly to him. One bit of history, which is not mythical in the least, is that one morning a few days after the *denoument* Frederick came, deflated and actually apologetic, to say it was all a pipe dream because at one stage the construction could not be continued. That's when I really jumped out of my skin.

Now a stage further back in pre-prehistory. The two levels idea, with one consisting of operators on the first, owes something to Eddington's vector space and associated energy-momentum array. Eddington had to get numbers out of dimensionality, and we set about (algebraically as we then said) getting rid of all the geometry that could not be part of any such argument. We knew that Eddington's numerology was not right because he did not make that separation, but we went on the assumption that perhaps the numbers would turn up somehow. We envisaged statistical techniques with the Pask machine in which the imposition of a space/time distinction of the units would be reflected in the selection from the random input. Frederick imploded on the scene at this point and pre-prehistory came to a full stop. It is worth pointing out that the Pask machine, with its random input, was explicit about the dynamical nature of the algebra. Who or what worked the algebra if it represented the real world? Frederick never wanted to think about that; but I kept harping on about it and it was about that time when he took a quite new turn with his indistinguishables. I do not know if there was any connexion.

What we did know something about was the first level. We got from the quadratic group (see 'Concept of order') an argument which derived dimensionality from ideas of symmetry as distinct from intuitive spatial perception, and that gave us the shape of the bottom level in combinatoric form. In fact that structure is isomorphic with Frederick's bottom level if you change multiplication for addition. Our assumption from those days, that the peculiar symmetry of the quadratic group explained why space was 3D, has turned out to be a half truth, and the full story is much longer. However, like believing in God, it is more accurate to believe that than to believe nothing.

REPRESENTING MEASUREABLE NUMBERS IN THE HIERARCHY

Ted Bastin

Maesllwyn, Tan y Groes, Cardigan SA43 2JF
(edward@bastinsooo1.fsnet.co.uk)

Draft submitted 1

ABSTRACT

This paper accompanies "Interpreting the hierarchy: Level change and Lou Kauffman's construction of matrices" by Clive Kilmister. It deals with the physical setting that demanded these mathematical innovations. This very general approach requires us to make an initial identification of hierarchy numbers with a scale of mass. Attention is drawn to correspondences of this approach with that of Rowlands and Cullerne. Future work is forecast concerning the early appearance of Pythagoras' theorem in any digital construction of geometry.

1. Physical variables in a combinatorial approach as distinct from bounded interpretable numbers

We pursue the path of constructing combinatorially based dynamical quantities to replace those of classical mechanics, and the proof of the pudding will be that the new go over into the old as a special case. I draw attention to David McGoveran's conclusion (Bit String Physics: Noyes P.298) that all existing attempts at a discrete physics seem to him to be "contaminated with continuum ideas". How difficult it is to get people to react to statements like these: they seem to take them in a spirit of toleration and fail to realize that they must be either ridiculous or else appallingly right. Under no circumstances may one say "Hum: interesting".

The combinatorial physics approach makes its initial contacts with experiment by identifying whole numbers, or ratios of whole numbers derived from the initial whole numbers, with physically important dimensionless numbers. There is a huge conceptual step from here to getting a representation of physical variables as they are usually understood. I will quote again Tom Etter's characterization of this step as being from 'composition' to 'extension' in order to exhibit its prevailing importance. A related recognition of this step is Kauffman's discussion of what he indeed calls 'steps', but more on that anon. On a question of

suitable terms, I contrast the method where we interpret whole numbers with that in which we have to advance to *the continuum,* or at least to a dense set of unspecified rationals, before we can get to the physical variable: that begs the question of what the continuum really is. It is better to say that we need to contemplate an *unbounded set* of interpretable numbers.

The first things that have a physical identification are the coupling constants. However they do not give us a range of values. To make that leap we think the only way forward is to envisage a continual flux of elements. This involves a corresponding continual jumping between levels. Then the process allows the alteration of the vector at the more complex level, and so the selection of a different discriminately closed subset. Then one may revert to the simpler level. This to-and-fro could go on for ever, so that we have escaped the boundedness.

2. The matrix picture and permutations, Kauffman's treatment.

If we are to use the matrix algebra in the combinatorial hierarchy to escape boundedness then we must look again at the function of the matrix transform in that mathematics. Trouble appears because the usual introduction of matrices in the hierarchy algebra is by creating a new entity which replaces a dcs as a new entity in its own right, forming what is then called a new level with vectors of squared length. In the first place it has been assumed that a matrix will specify a dcs to represent an area of physical activity within whatever the whole structure represents –the structure of a particle for example. This activity or flux goes on without intervention by the matrix However the matrix has a second function which needs more scrutiny. In essence a *permutation* of any vector at will is what is required. These two functions or roles for the matrix are not obviously compatible. We have often spoken of 'working at a given level' in recognition of the fact that if the two things the matrix does are separated in real time then there can be no conflict. Of course the apparent anthropomorphism of this language is only to be brief. Everything has to happen automatically, though we do suppose that there will be ways in which an experimenter can learn to alter the balance of the choices at the different levels in order to interact with the world.

The matrix thus has the function of changing one dcs for another by permutation of is own elements. The essential step at this point is to realize that all the elements of the vector must be treated exactly the same. We call such treatment 'indifferent'. In his paper, Kilmister set out an algebraic structure meant to handle the dynamical aspect. In earlier discussions of the hierarchy algebra we have failed to realize that we could not just import the matrix as Parker-Rhodes did, but have to explain what it is about it that plays the dynamical role. Kilmister used an

ingenious reconstruction of matrices in terms of permutations due to Kauffman. We are trying to find what operator will preserve the essential character of a vector or ordered string while permitting the flux of its elements. This operator must work elementwise on the elements of the string to permute them and so we form the *intersection* of the things and we may conveniently call this multiplication. For example the operator 0101 would permute 1101 into 0100. Now we have to incorporate the fact that all the permutations of the operators have to be *combined.*

The word *combine* has two component meanings –both essential. First, it means that the operation of every element must alter the result: secondly it means that the final result must be independent of the order in which the elementwise operations are carried out. But the converse is not true and both are needed. These two conditions satisfy the requirement of simultaneity and of the *indifferent* treatment of the elements. Clearly we need a new elementwise operation which sums up the effects of all the component elements. If we were working over the rationals the obvious thing would be to call this addition to give a sort of average, and then we could advance straight to defining the matrix. Working combinatorially things need more thinking out. In fact at any given stage there are only two results –existence and non-existence represented by 1 and 0, the combination will have to be statistical in character. These considerations restrict us to symmetric difference denoted by the + sign. Thus we find discrimination appearing again, and in fact we have arrived at matrices composed of 1 and 0. It will be important in applying the construction to dynamic operations to realize that what we do is conditional on our observing all these considerations which led us to the 'matrix' form.

3. What this replacement of the old matrices imposes on the interaction of levels.

One advantage of the new approach to the matrix formulation is that it makes the notion of level much less artificial since it is consequent on the need for changing permutations which we cannot imagine our being without.

One general consequence follows from the permutation construction which I shall come back to but which I mention in a bare form now. The possibility of working at more than one level is essential for getting from bounded physical quantities to unbounded ones. One may use a matrix to specify a given permutation and thus specify a physical change. Having made the change one may revert to the level at which one considers the elements of the matrix and make a change there. It is necessary to have succession of these stages to get enough logical

richness to represent unbounded change among elements which are in their nature bounded until some new principle is brought in.

It is difficult to avoid slipping into anthropomorphic language as I have just done in speaking as though we had to decide what level we are working on at any given moment. Of course the reality is that the changes are imposed by the input to the whole system so that at a given moment there is a flux of activity in all the DCSS at one level and then at another moment a switch is made which controls the persistence or non-persistence of this activity. The changes are in real time but we do not control them. You may remember that I made a great fuss about recognizing that we must contemplate other systems like our own in the 'vasty deep' so that input to the system could take place at all levels. This let me out from solipsism.

In the Kauffman-Kilmister view of the origin of the matrix the stability imposed of the flux of each system is imposed by a sort of averaging over an extended period of activity which requires a new logical operation. Kilmister shows that this may be taken to be the discrimination operation because of the strong symmetry thus introduced. He would like to be able to say that it *has to be*, but the sufficient condition is good enough to go on for me..

A direct consequence of the bounded/unbounded switch is that some numerical measures associated with the former are squared in the latter. This follows from the Kauffman definition of multiplication. I mention now that we shall find in this squaring the origin of the squaring of momentum in the energy-momentum relationship of special relativity. I mention such a far consequence at this point merely to foreshow that a lot of detail of importance will hinge on what is being said.

4. The new dynamical concepts

In getting to what I call the physical variable as distinct from bounded discrete entities one has to advance in stages. The first things that have a physical identification are the coupling constants. However they do not give us a range of values. Even when we have advanced to having a range of values the main step towards being unbounded in the way the classical dynamical concepts are, has still to be taken To understand this progression of thought one must see the inevitability of reinventing *all* of the classical concepts completely. This requirement is very strange since we are so used to concepts which are defined in terms of spatial separation. So strange is it that it is difficult to get the reader to react to the proposal for a complete replacement. People usually behave as if they haven't heard what you are saying or as though it is part of the inessential gas that people must be allowed for the sake of politeness.

There is another major change in outlook that we have to make which is hardly separable from the reinvention of classical concepts that goes with our view of the physical variable. One naturally needs to have a physical picture of the particles of high energy physics: Now what could we mean by a physical picture which does not presuppose the spatial continuum? One might assume that thinking of this sort would take the form of abstract groups and such algebraic devices, which in themselves could not have physical consequences. Do we mean that? Not bloody likely. In the foregoing discussion I have spoken as though changes in –say- groups are at once things going on in the world. They happen in real time. This is the brutal face of the process view. Thus I contemplate atomic structures built up of combinatorial motions that are in every way real, and underlie and indeed dictate the dimensional shape of the spatialized phenomenon. I think of an analogy which is helpful if it is not pushed too far. The working of Newton's pendulum –the way when you let two balls impact, upon which exactly two detach themselves at the other side, for example- has a purely combinatoric explanation even though it deals in the kinetics of matter. Someone who should have known much better once told me it was the only way in which the principles of conservation of energy and conservation of momentum could be reconciled. In fact of course the principles of dynamics do not enter at all.

One may say 'why exclude space while having an all-too-evident time'. The answer is that time and space are different. The difference is obscure here because what we are calling time is simply succession, not as the full physical coordinate of –say- relativity. We can stop short of that extension, but we cannot do without a principle of succession.

5. Given mass/charge the first variable has to be momentum

Our first physical identification has to be mass because it appears as a factor in the fine structure constant. I described last year how for Rowlands and Cullerne the construction of mass and charge must go in tandem, so that in the step-by-step filling up of the scale from zero to 137 we were defining a primitive additive quantity which might equally be termed mass or charge since no structure has yet appeared for us to distinguish them. Mass in fact appears through the factor h which has the dimensions of action ML/T. Rowlands and Cullerne have a theory of 'charge exclusion' which is how integral numbers enter physics. They apply this principle to introduce combined charge and mass. I quote from the first two paragraphs of P.10 of "Investigation of the Higgs mechanism". "The Higgs mechanism was introduced earlier to account for the spectrum of particle masses. We have previously shown [R&C 200a] that the breaking of symmetry which produces mass is largely a

result of the production of zero states of charge, mass and charge acting as an invariant in the same manner as space and time. Zero charges represent complete coupling to the Higgs field, non-zero charges represent a reduction of the vacuum state to less vacuum. The complete coupling to a zero charge may be equivalent in energy to one unit of m_e/α."

Here we have a very strong tie-up with the hierarchy picture which may be conveyed in a few points:

1. The electromagnetic level provides a discrete scale of magnitudes which has the (uncorrected) reciprocal of the fine-structure constant as its maximum -137. This progression should be identifiable with R&C;s charge exclusion.
2. Since both e and m appear in the dimensionless form of this constant ('m' via h) we are committed to a mutual generation of m and e. We do not get one without the other. This accords with the position of R&C.
3. The correspondence between the two theories at this point is more evident if we recall that in both cases the mechanism comes from the group transformations directly.
4. For us, this is the first appearance of a quantized magnitude with physical interpretation, and everything proceeds from here. For R&C the situation is more complicated since they take over the concepts of mass and charge from continuum physics
5. R&C would seem to have a basis for quantization itself starting here. Here 'quantization' means associating integers, and subsequently perhaps rationals with the numbers derived from the algebra of the two-group picture. Such a construction is already entailed when we impose some finite and fixed value to Planck's constant. You could say that they are greatly underplaying their hand.

We need an alternative to the classical particle mechanics in which one presumes an understanding of the massive particle and then defines momentum in terms of the spatial motion of that particle. As we have the picture at present, one still is using an intuitive continuum idea to place points or values of momentum/mass on a scale. The Dedekind section still seems to incorporate this appeal. I am reminded of our contention from long ago that Brouwer's 'continuum' with the freely proceeding sequences is what is needed. Belatedly, we are now utilizing this in the sense that we must always answer the question "where do we site a new value" by going back along the particular history of constructions in the hierarchy. Thus we avoid arguing in a circle. I think this view becomes more natural in the light of the permutations, but I do not quite know how to put that. If I am right then it is a bit misleading to write as though the steps up to the fine-structure constant to provide mass/charge, are discrete

stages between which we interpolate new values. That is not to say that the steps do not have an essential role, but that role arises through their being the particular form that is needed to construct sequences. That is to say there has to be *something*: just having one step would give a trivial structure.

Mass and momentum are inseparable in the sense that when we put in place the means of getting a variable we created the momentum concept (and I suppose other less primitive concepts) in the act of doing so. I am trying to get away from the idea that we first defined extension and then populated it with momentum and the rest.

6. Pythagoras

The origin of the squared form in the relation of mass to momentum in conventional theory must come down one way or another to Pythagoras' theorem. If we find a different origin for that squared form then we must be claiming to derive Pythagoras as a condition for setting up geometrical space. That is perhaps a bit sensational.

The mystery is resolved a bit when we consider that, working the way we are, the first steps in geometry are indissolubly linked with special relativity. An undetermined factor enters in the connection of m to p and we write this c. Then we are free to follow the history books in reverse. Usually, we think we have Euclidean space that we tinker with to make Lorentz contractions appear and so get a finite signal velocity. Now we can take the constant c as predetermined, but regard the interpretation of it as a maximal velocity as the condition under which we can have Euclidean space as an approximation at small velocities. This change in viewpoint must have important consequences for cosmology – for example over the interpretation of the red shift as a recession velocity- but I shall not pursue this matter just now.

Of course momentum and spatial velocity are not the same and the clue to the necessary separation of them lies in our having treated the quantity h in the fine structure constant as simply a scaling factor (having dealt with m and e properly) when it has actually the dimension of action. Perhaps we were wrong in picking on linear momentum when we should not have allowed the decision on that interpretation to go by default, but been explicit in choosing angular momentum. That would give angular momentum the pre-eminence it has in quantum physics but the connection with relativity becomes unclear. However the whole argument is useful in showing the basic thinking that is inevitable in our reconstruction of the dynamical variables.

INTERPRETING THE HIERARCHY: LEVEL CHANGE AND LOU KAUFFMAN'S CONSTRUCTION OF MATRICES

CLIVE KILMISTER

Red Tiles Cottage

High Street

Barcombe

Lewes BN8 5DH, UK

I describe work done in conjunction with Ted Bastin which shows that two algebraic notions of Lou Kauffman allow one to progress from the static Combinatorial Hierarchy to a more dynamical picture.

This paper is about the Combinatorial Hierarchy (CH) as formulated originally by Frederick Parker-Rhodes and described in (1). It is not concerned with non-commutative discrimination. My flagging interest in that has been rekindled by reading Doug Matzke's thesis but that is a matter for next year and will be a major task. Now I seek to throw light on the dual role of matrix transformations, both in level change and in describing physical change. The paper is in two parts. The first part explores features of discrimination which should have been looked at in Combinatorial Physics (CP), (2). The second part tackles the main problem of level change and dynamics. The paper hinges on two pieces of information from Lou Kauffman.

PART 1. DISCRIMINATION

In the process view, which I have urged as the only way to understand the CH, new entities continually come into play. When a putative new entity comes into play it has to be "checked" to see if it is a genuinely new one or a repetition of an earlier one. This "checking" is done by a discrimination operation, D. If a, b are two elements in question, Dab has the value of a signal, Dab = 0 (in the notation which takes 0 as the only signal) when a, b are two occurrences of the same entity. If Dab = c $\neq$ 0, then c is a new element. This new element is the bit of information that a, b have been discriminated and found to be different. In CP this is expressed in terms

of an ordered set of symbols, L = [1, 2, 3, ...] and the rule ("Conway's rule") that Dab is to be assigned the least symbol which has not been used already to express Da'b or Dab' where a', b' are any symbols less than a, b respectively. I want to emphasise that Conway's rule is simply a statement of economical organisation of symbols, not a physical input. None the less, it easily gives the commutative rule Dab = Dba and the associative rule DDabc = DaDbc. Because of commutativity one can rewrite D as an addition operation Dab = a + b and then the associativity means that no brackets are needed in a + b + c.

This was regarded in CP as a useful mathematical fact, which it was. We paid less attention to the fact that the system was showing us something. No meaning was given to a + b + c except in terms of successive discriminations of pairs of elements, e.g. as a + (b + c).

Now to remind you of much earlier work I did with Ted Bastin. In our Concept of Order papers, specifically (3), we considered a set E of elements and a set F of relations, some dyadic, some triadic and so on. In that paper all the relations of F were constructed from one basic triadic relation by existential substitution[1] *. We called the elements simultaneous if and only if, whenever any subset of E with k elements satisfied a k-adic relation of F, then so did any permutation of the k elements. Thus, if Rabc, then also Racb, Rcab etc. We also introduced a stronger restriction called similarity of position. The elements of E had similarity of position if, whenever a k-adic subset of E satisfied a k-adic relation in F, so did any other k-adic subset. Thus, if Rabb, then also Rbaa, Rbcc and so on. We found that , if E were a set of simultaneous elements, then Rabc could be written as ab = c and E was an abelian group in which every element (except the unit) was of order 2. By a well-known result, E was then a direct product of cyclic groups of order 2, $E \simeq C_2 \times C_2 \times \ldots \times C_2$. The stronger condition of similarity of position limited E to $C_2 \times C_2 = S$, the quadratic group. The isomorphism

---

* Superscripts refer to notes at the end of the paper.

of these groups to the later ones of the CH was, of course, no coincidence.

It would have been better if we had interchanged the two names and called the weaker similarity of position. After nearly half a century it is too late to change and so I remark that the elements in a + b + c have the property of simultaneity. That is, if we define a relation

$$R_3abc \leftrightarrow a + b + c = 0,$$

then if R_3abc holds, so does R_3xyz where x, y, z is any permutation of a, b, c. Moreover, if R_3abc does not hold, so that $a + b + c = d$, then $a + b + c + d = 0$, which we could write as

$$R_4abcd \leftrightarrow a + b + c + d = 0.$$

And if R_4abcd does not hold, then

$$a + b + c + d + e = 0 \leftrightarrow R_5abcde,$$

and so on[2]. The process picture and the need for discrimination necessitate this unexpected set of relations. I want to enquire what these relations mean.

The first two relations, R_2 and R_3 are easy. R_2ab means $a + b = 0$, that is, $a = b$, the original discrimination in a different notation. Then R_3abc means $a + b = c$ (and the other five versions of this got by permuting) and so is simply the statement that [a, b, c] is a discriminately closed subset (dcs). But the higher order relations have not hitherto been noticed and so have no names. For example, one could say that R_4abcd means that d is the operation of determining whether [a, b, c] is a dcs and concluding that it is not; but this way of putting it obscures the simultaneity of a, b, c, d. I need a new name. For reasons which will become clear at the end, I shall call [a, b, c, ... z] a permanent set if and only if $a + b + \ldots + z = 0$, i.e. $R_k abc..z$. It may be necessary to refine this concept a little. Call a set K a primitive permanent set (pps) if it is permanent but no proper subset of K is permanent. Thus any dcs is also a permanent set but only the quadratic group S is a pps. Evidently there are ppss of any size, unlike dcss. The concept of a permanent set is related to the usual one of a linearly

dependent one (over $\mathbb{Z}_2$) but there are differences. I shall show later that the notion of permanence gives some physical meaning to these relations.

PART 2. LEVEL CHANGE

I return to the process view of the CH. Parker-Rhodes chose to exhibit the CH in terms of bit-strings, that is, vectors over $\mathbb{Z}_2$. Then discrimination is just addition over $\mathbb{Z}_2$. In CP we showed that this is almost general; "almost" because Parker-Rhodes bit-strings were of definite length(dimension) whereas the elements thrown up by the process were not. Ted prefers to retain the bit-string picture, none the less; I think it emphasises the vector-space picture too much and prefer another picture set out here. As so often with my thinking, it is through care with notation that new concepts arise. The usefulness of the bit-string picture is that, if an element is (say) $\begin{bmatrix}1\\0\\1\\1\end{bmatrix}$, this can be interpreted as a physical entity for which certain attributes are present (viz.,the first, third and fourth). My preference is for the "row" notation of CP. A <u>row</u> is simply defined as a ordered string, pqr... of elements of L such that $p < q < r < \ldots$. Then the vector above would be written as a row 134. The concept of a row is a small but essential generalisation of that of a vector. If one had to, one could say that the row stands for the infinity of vectors $\begin{bmatrix}1\\0\\1\\1\\x\end{bmatrix}$, where x is a column of zeros of any length. In this way the dimensionality is not fixed and the vector space is not suggested.

What Parker-Rhodes emphasised was that, given two elements, one automatic automatically built up in succession aggregates of 3, 10, 137 and 10^{39}. There is a certain static character about this. It just says: here is a hierarchy. It reminds me a little of the old books on electromagnetic theory which reached Maxwell's equations in the last chapter and then just admired them. To progress further with the CH the concept of <u>change</u> has to be introduced. Our general idea is a continual flux of elements represented by bit-strings or rows, regarded as statements of the presence or absence of attributes. The first step to dynamics is to allow the situation in which an attribute

ceases to be present. Thus the row 134 above might become 13. A dynamical operator, call it O, is needed to describe this. This O must operate independently on each bit of the bit-string. If the bit-string bit is 0 it must remain so. If it is 1, then it may remain so unless it is one of the bits to become zero. These conditions make it clear that one can use element-wise multiplication (over $\mathbb{Z}_2$) to construct O. One can define the product of two bit-strings $\begin{bmatrix} a \\ b \\ \cdot \end{bmatrix}$, $\begin{bmatrix} p \\ q \\ \cdot \end{bmatrix}$ as the bitstring $\begin{bmatrix} ap \\ bq \\ \cdot \end{bmatrix}$. Notice that non-element-wise products like aq do not occur. This version of "multiplication of vectors" was suggested to us in a letter from Lou Kauffman. Then the stated removal of one attribute would be of the form $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ as required. And there is an obvious set of basis elements O_1, O_2, O_3,... where

$$O_k^2 = O_k \ , \quad O_i O_j = 0 \ (i \neq j)$$

so that any sum of O_k's is also idempotent.

It was simplest to carry out that discussion in terms of bit-strings. I now verify that it carries over to the slightly more general rows. Recall that in CP addition of rows was defined in this way:

(i) if X is <u>any</u> sequence of the symbols of L, then $\varrho(X)$, the "row of X," is got by (a) re-arranging the symbols of X in order, (b) deleting any adjacent pair.

(ii). If x, y are rows, write $x \cdot y$ for the sequence got by juxtaposing x and y. Then define

$$x + y = \varrho(x \cdot y).$$

How then to define multiplication of rows? A moment's thought gives

$$xy = \varrho(x \cap y),$$

where $x \cap y$ is the ordinary intersection of the two rows considered as sets of symbols.[3]

It is time to confront a serious problem. We agreed that 135 (say) is a row which declares that certain attributes are present. It is tempting, but it would be wrong at this primitive stage, to try to identify these

attributes by, for example, comparing them with known quantum numbers in the standard model. The authors of (1) made this mistake. The reason that it is a mistake is this: all that has been generated at this stage is a string of descriptors. It will be some way down the line before the system is complex enough to generate a descriptor with enough relational properties to be identified, e.g. with spin. But thinking about this mistake shows something about the descriptors at this stage. They must be treated all on the same footing. This phrase is somewhat ambiguous and will need clarifying in the future. The point is that to distinguish one of them and say that it represents spin, say, would be to interfere with the system from outside, which the process view prohibits. So our system must treat all the 10 permutations of 135 indifferently. That is, as well as 135, 123, 124, 125, 134, 145, 234, 235, 245 and 345 must all come in. The row notation is useful here; all permutations of a row simply means all rows of the same order (the order of a row being its length). This means that I must extend the notion of a dynamical operator. Instead of considering a row multiplying 135, I define the generalised dynamical operator, O, acting on the set s_5 of ten elements,

$$s_5 = [135,\ 123, \ldots . 345].$$

Then $O(s_5)$ is also a set of ten elements:

$$O(s_5) = [v_1(135),\ v_2(123),\ \ldots\ v_{10}(345)]$$

where the ten v_k are left-multipliers like 123 in the example above. It is now no longer true that $O^2 = O$.

My next problem is to isolate states of affairs which preserve some permanence in the flux of change described by O. This permanence must be defined by some relation R satisfied by the ten elements of $O(s_5)$. But we have to treat the ten elements of $O(s_5)$ on the same footing, so that if some relation holds between the ten, they must have simultaneity under it, in the sense described earlier. This is the reason for my discussion in PART 1. One set of such relations is the set R_2, R_3, ... defined there. I cannot claim that these relations are the only possible ones with simultaneity.

It does seem likely that they are the only ones constructible from discrimination alone. But that is a very weak result. I would like to know a minimum set of assumptions that would finish with these relations, and then I would be able to judge whether these assumptions were physically determined. For the present I assume that the appropriate set of relations are these and so a condition for permanence in the flux is

$$R_{10}(s_5) \leftrightarrow v_1(135) + v_2(123) + \ldots + v_{10}(345) = 0.$$

This leads to my using the second piece of information from Lou Kauffman. He puts it this way: let u be any vector of dimension n (over any field F) and define $p_A(u)$, A = 1, 2, ...,n!, as the vectors got from u by permuting its components. Then $\sum_A v^A p_A(u)$ is just the general definition of a matrix operator, Vu. Taking this result with $F = \mathbb{Z}_2$ shows that the condition for permanence in the flux is exactly that used in level change, when a dcs at one level is described at the next higher level by the matrix which annihilates it and nothing else. Parker-Rhodes preferred the condition Wu = u rather than Vu = 0 but one is transformed into the other by W = V + I, where I is the identity.

To summarise where I have reached so far: firstly, at the mere technical level, Lou's formula gives me the corresponding small generalisation of the matrix operator appropriate to rows. Secondly, it appears that such operators have a dual role. As well as their well-known level-change role, they are a first step towards dynamical change, such change coming when Vu = w ≠ 0 In fact, Ted was always clear that level change operators played this dual role but I was not.

NOTES

1. That is, from (say) R_3abc one constructs an R_4 by

$$R_4abef \leftrightarrow (\exists c)R_3abc \;\&\; R_3cef,$$

and so on.

2. One may note that

$$R_n xay \leftrightarrow [a = \neg R_{n-1} xy]$$

where x, y are (possibly null) sets of elements and a is a single element.

This way of putting it emphasises that the system is not restricted by a simplistic theory of types. This may be a danger signal for later on.

3. It is tempting to say this fits in nicely with set theory; isn't x + y just $\varrho(x \cup y)$?Unfortunately not, because (123) + (134) = ϱ(123134) = ϱ(112334) = 24, whereas (123) $\cup$ (134) = (1234). In fact, if one wants to go down that road, one could write

$$x + y = \varrho[x \cup y - x \cap y]$$

REFERENCES

(1) "On the Physical Interpretation & the Mathematical Structure of the Combinatorial Hierarchy" by T.Bastin, H.P.Noyes, J.Amson & C.Kilmister. Int.J.Theor.Phys. 18, 445-488, 1979.

(2) Combinatorial Physics by T.Bastin & C.Kilmister, World Scientific,1995.

(3) "Concept of Order,I. The space-time structure" by E.Bastin & C.Kilmister. Proc.Camb.Phil.Soc. 50, 278-286, 1954.

Representations of a Fundamental Theory

Peter Rowlands

IQ Group and Science Communication Unit, Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford Street, P.O. Box 147, Liverpool, L69 7ZE, UK. e-mail: p.rowlands@liverpool.ac.uk

Abstract. A series of representations of the fundamental theory that physics has a symmetrical or dualistic structure is used to show that both physical and mathematical ideas result from an attempt to maintain a zero total content in any fundamental conceptual scheme for explaining nature.

INTRODUCTION

The fundamental theory to be represented is the idea that physics has its origins in a symmetrical structure which preserves its conceptually zero content. It has become popular in recent years to suppose that the total energy in the universe, counting gravitational energy as negative, may be zero, and that the entire universe may have emerged out of pure nothingness through something like a quantum fluctuation. Perhaps also, if we include the vacuum state, there is some kind of ultimate balance between matter and antimatter. It is not unusual, in fact, to read statements like the one by the well-known chemist and science writer Peter Atkins [1994], who has said of physical matter that 'the seemingly something is elegantly reorganized nothing, and ... the net content of the universe is ... nothing'. However, it is not just matter and the universe that appear to be nothing, but the entire conceptual scheme of which these are merely components (*nihil ex nihil fit*). Mathematically, only zero is absolutely unique, and this leads one to suppose that an absolute or universal theory must in a sense be, ultimately, a theory about the meaning of nothingness. The idea of conceptual nothingness has been proposed on many occasions by the author as the basis of physics [Rowlands, 1983, 1991, 1999, 2001], in the sense that the perfect symmetry between the only truly fundamental parameters in nature is exactly of this sort, even in algebraic terms. In view of the intrinsically mathematical nature of physical quantities, and the probability that mathematics and physics have the same conceptual origin, it also seems to make sense to describe mathematics in terms of the same totally zero structure. The question then becomes: how do we get something from nothing? Such a question is unlikely to be answered on the basis of an

appeal to pure first principles without some initial empirical investigation into the way that physics and mathematics appear to be structured at their foundations. A convenient starting-point is the evidence of duality, or, rather, *something which leads ultimately to the creation of this concept*, as a fundamental principle.

DUALITY AND ITS GROUP STRUCTURE

Duality is fundamental to physics and mathematics. All operations are dual operations. All objects are dual objects. All physical and mathematical theories are dual theories. As Nicholas Young writes in a well-known textbook [1988], 'the idea of duality pervades mathematics', while duality has, from the beginning, been the basis of the author's theories of mathematical physics and philosophy of science [Rowlands, 1983, 1991, 1999, 2001]. In effect, we can't define something without defining also what it is not. Alternatively, we can't characterize 'nature' or 'reality', even to the extent of saying whether it has an independent existence (is ontological) or is a product of our perception (is epistemological). So, the concept even has a fundamental philosophical manifestation. Any attempt to characterize it in one aspect will automatically lead to our discovery of the 'opposite' characterization in another. Every 'probe' will meet with an opposing 'response'.

Why is this? It seems that physics and mathematics, as we have supposed, are attempts at creating something from nothing. Although we assert that we have given nothing a character or aspect, it is, in fact, still nothing, if we take the totality of probe and response. An obvious case is provided by the conservation of linear momentum in an explosion problem; gain of positive momentum in one direction has to be countered by an equal gain of negative momentum in a direction which is precisely opposite. The totality remains zero. The example is a very familiar one, and it is usually treated as an illustration of a very fundamental law of physics, with an independent existence. The law is, in fact, however, just one illustration of a very much more fundamental principle of duality in nature, which can be seen as the ultimate origin of both physics and mathematics. A 'theory of everything' needs first to be a 'theory of nothing'.

How, then, does this principle of duality operate? The answer has to be: in the simplest way possible. It is not possible to imagine any duality simpler than that provided by the C_2 group. We could describe it, in mathematical terms, by the use of the elements 1 and –1, but our starting concept must, in fact, be even simpler than that, and cannot yet assume the discrete numbering associated with the term 'dual'. The concept of 1 may appear to be simple, but it is in reality loaded with information about

the meaning of discreteness, in addition to ordinality, which doesn't appear at the most fundamental level. A simpler opposition is that between + and – applied to the unspecified entities which are generically described as the reals. A mathematical-computational approach to this is given by the author, in a paper written with Bernard Diaz [Rowlands and Diaz, 2002].

In effect, the simplest possible thing other than 0 that we can imagine is $\mathcal{R}$, or can be described as $\mathcal{R}$, where $\mathcal{R}$ is a totally unspecified or undifferentiated entity, and its automatic negation or 'conjugation' is the thing we describe as $-\mathcal{R}$. In principle, as soon as we have $\mathcal{R}$, we have no option but to take $-\mathcal{R}$ as well, in order to maintain the zero totality. That is, defining $\mathcal{R}$, at all, automatically creates what we will eventually call a 'dual' system. Ultimately, we will find that this is equivalent to requiring 1 + 1 = 2, and generating the Peano idea of 'successor', and a natural (binary) numbering system, which avoids the Gödel problem through a zero totality, but this will require the generation of a concept of discreteness which is not a direct feature of $\mathcal{R}$. We cannot, at this stage, even take $\mathcal{R} - (-\mathcal{R})$ to be, say, $2\mathcal{R}$, until we have defined 2, and the concept of number generally, to exist. The existence of + and – signs can thus be taken as an expression of ordinality, but not yet of a discrete ordinality (as with the Dedekind 'cut', which, despite its name, is a definition of ordinality without a prior assumption of discreteness), and the expression of this process as *duality*, or the group structure as C_2, cannot yet be taken as explicit.

We also have no option but to relate $-\mathcal{R}$ to $\mathcal{R}$ in some way other than defining their totality as 0, and the identity $-\mathcal{R} \times -\mathcal{R} = \mathcal{R} \times \mathcal{R}$ then becomes deeply significant in establishing that the relation between these elements is a *group* relationship, and that the 'multiplication' and 'squaring' of elements, in addition to identity and inversion, are operations which are fundamental to the principle that we will ultimately describe as 'duality', when we have introduced discreteness. Of course, without some concept of enumeration, and no way of identifying $\mathcal{R}$ more exactly, $-\mathcal{R} \times -\mathcal{R} = \mathcal{R} \times \mathcal{R}$ is describable only by the generic $-(-\mathcal{R})$ or $\mathcal{R}$, there is no defined concept of 'unit', and the × operation is not yet identifiable as 'multiplication'.

But, suppose now that we require a counter concept, even for the $\mathcal{R}$ and $-\mathcal{R}$ category, that is, we require a system which avoids 'privileging' these, and privileging one with respect to the other. We may then suppose that conjugate terms must exist which allow us to generate $-\mathcal{R}$ in the same way as we generate $\mathcal{R}$ from $-\mathcal{R}$ (in addition to its generation from $\mathcal{R}$). In mathematical terms, we describe these as members of the complex set, and each must have its own conjugate. Symbolically, we represent the

new terms as $\not{C}$ and $-\not{C}$. However, the new category $\not{C}$ remains undefined in respect to the real category, and has no ordinal relation to it. Consequently, there are infinitely possible or indefinitely possible systems and combinations that are represented by this symbol.

However, when we investigate the combinations of possible $\not{C}$ terms, we find a distinct separation between the infinitely possible combinations leading to the original real category $\mathcal{R}$, and the very definite *noncommutative* ones leading to the conjugate $-\mathcal{R}$, where only 'one' independent $\not{C}$-type concept (say $\not{C}'$) is associated with each conceivable $\not{C}$. Thus, we find that the former are infinitely extendible, while the latter are cyclic or enclosed. It is at this point that we can introduce discreteness, and the concept of 'unity', into mathematics. By choosing the default position of assuming indistinguishability between the $\not{C}$s in every conceivable respect (i.e. none is 'privileged'), we can create a regular ordinal sequence, which, although arbitrary in principle, becomes a series of integral binary enumerations, which can also be applied to ordinality in the real categories. The consequent minimising of variation within the generating process additionally allows us to retain (and even create) the powerful notion of group structure. We can represent the generality of the process in the form:

$\mathcal{R}$	undefined
$\mathcal{R}, -\mathcal{R}$	conjugation
$\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}$	complexification
$\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}, \not{C}', -\not{C}', \not{C}\not{C}', -\not{C}\not{C}'$	dimensionalization
$\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}, \not{C}', -\not{C}', \not{C}\not{C}', -\not{C}\not{C}',$ $\not{C}'', -\not{C}'', \not{C}\not{C}'', -\not{C}\not{C}'', \not{C}'\not{C}'', -\not{C}'\not{C}'',$ $\not{C}\not{C}'\not{C}'', -\not{C}\not{C}'\not{C}''$	repetition

The logical operations involved in the sequence can be expressed in a quasi-algebraic form though operations such as × and – are not limited to an algebraic interpretation until we create the concept of integral sequencing via the ordinal series of closed systems:

$$\mathcal{R} \times \mathcal{R} = -\mathcal{R} \times -\mathcal{R} = \mathcal{R}$$
$$\mathcal{R} \times -\mathcal{R} = -\mathcal{R} \times \mathcal{R} = -\mathcal{R}$$
$$\mathcal{R} \times \not{C} = \not{C} \times \mathcal{R} = \not{C}$$
$$\not{C} \times \not{C} = -\not{C} \times -\not{C} = -\mathcal{R}$$
$$\not{C} \times -\not{C} = -\not{C} \times \not{C} = \mathcal{R}$$
$$\not{C}' \times \not{C}' = -\not{C}' \times -\not{C}' = -\mathcal{R}$$

$\not{C}\not{C}' \times \not{C}\not{C}' = -\not{C}\not{C}' \times -\not{C}\not{C}' = -\mathcal{R}$ closed (anticommutative)
$\not{C}\not{C}'' \times \not{C}\not{C}'' = -\not{C}\not{C}'' \times -\not{C}\not{C}'' = \mathcal{R}$ unlimited (commutative)

The character sets effectively represent all those, including $\mathcal{R}$, $-\mathcal{R}$ which are generated by operating on themselves:

$(\mathcal{R}) \times (\mathcal{R}) = (\mathcal{R})$
$(\mathcal{R}, -\mathcal{R}) \times (\mathcal{R}, -\mathcal{R}) = (\mathcal{R}, -\mathcal{R})$
$(\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}) \times (\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}) = (\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C})$
$(\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}, \not{C}', -\not{C}', \not{C}\not{C}', -\not{C}\not{C}') \times (\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}, \not{C}', -\not{C}', \not{C}\not{C}', -\not{C}\not{C}')$
$= (\mathcal{R}, -\mathcal{R}, \not{C}, -\not{C}, \not{C}', -\not{C}', \not{C}\not{C}', -\not{C}\not{C}')$, etc.

The closed sets are those which introduce discreteness through anticommutativity.

We are now in a position to extend the argument using the integral sequence we have created. Beginning with the C_2 group, which can now be represented by 1 and −1, a dual system will extend this to four elements, producing an equivalent to $C_2 \times C_2$, and we choose the only way of extending a group including 1 and −1 to encompass four elements, by making the unknown elements (hitherto represented by the generic $\not{C}$ and $-\not{C}$) acquire the characters that we describe by the algebraic symbols i and $-i$. The group of 1, −1, i, $-i$ is not, of course, $C_2 \times C_2$, or D_2, but C_4. However, it contains the same *information* as $C_2 \times C_2$, for we can write this information in the form of the complex ordered pairs: 1, i; 1, $-i$; −1, i; −1, $-i$, which *is* of the form $C_2 \times C_2$, and is the only domain in which $\pm i$ can exist.

If we are now required to dual the C_4 group, the most efficient and ordinally-structured way of retaining elements equivalent to 1, −1, i, $-i$ in an extended group of order eight, is by supposing that we can expand i, $-i$ into the necessarily *cyclic* and noncommutative operators $\boldsymbol{i}$, $-\boldsymbol{i}$, $\boldsymbol{j}$, $-\boldsymbol{j}$, $\boldsymbol{k}$, $-\boldsymbol{k}$, which we describe as quaternions. The definition of the quaternion group Q_8, with elements 1, −1, $\boldsymbol{i}$, $-\boldsymbol{i}$, $\boldsymbol{j}$, $-\boldsymbol{j}$, $\boldsymbol{k}$, $-\boldsymbol{k}$, is simply a statement of the fact that the complex C_4 group has been dualistically extended on the basis that $\boldsymbol{ij}$ (= $\boldsymbol{k}$) has the same kind of properties as $\boldsymbol{i}$ and $\boldsymbol{j}$, with $(\boldsymbol{ij})(\boldsymbol{ij}) = -1$. Again, we can represent the same information by a C_2 multiplication, using a group of the form $C_2 \times C_2 \times C_2$. The cyclic nature of the quaternions is significant here, because the eight possible $(C_2 \times C_2 \times C_2)$ combinations of $\pm \boldsymbol{i}$, $\pm \boldsymbol{j}$, $\pm \boldsymbol{k}$ become sufficient to generate the entire information produced by the elements of Q_8. In effect, describing a set of operators, such as $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$, as 'cyclic' means reducing the amount of independent information they contain by a factor 2, because $\boldsymbol{k}$, for

example, arises purely from the product ij. It could even be argued that the necessity of maintaining the equivalence of the Q_8 and $C_2 \times C_2 \times C_2$ representations is the determining factor in making the quaternion operators cyclic. In addition, the cyclicity prevents the definition of further complex terms, such as I, where $(iI)(iI) = -1$, though there are an unlimited number of I terms such that $(iI)(iI) = 1$.

The process can be continued further using terms of this kind. We dual Q_8 by complexifying it to the complex quaternion or multivariate 'vector' group 1, –1, i, $-i$, $\mathbf{i}$, $-\mathbf{i}$, j, $-j$, k, $-k$, ii, $-ii$, ij, $-ij$, ik, $-ik$, of order 16, which has a related $C_2 \times C_2 \times C_2 \times C_2$ formulation, and which may also be written 1, –1, i, $-i$, $i\mathbf{i}$, $i\mathbf{i}$, $i\mathbf{j}$, $-i\mathbf{j}$, $i\mathbf{k}$, $-i\mathbf{k}$, $\mathbf{i}$, $-\mathbf{i}$, $\mathbf{j}$, $-\mathbf{j}$, $\mathbf{k}$, $-\mathbf{k}$, where a complex quaternion, such as ii becomes the equivalent of the multivariate vector $\mathbf{i}$ (see Appendix I). (It is significant, here, that a possible alternative dualling of quaternions to octonions, with sixteen components, would fail to maintain the group structure, as octonions are nonassociative.) We then expand the complex terms to a three-dimensional status, to produce a double quaternion group, say 1, –1, I, $-I$, J, $-J$, K, $-K$, i, $-i$, j, $-j$, k, $-k$, of order 32, which has a related $C_2 \times C_2 \times C_2 \times C_2 \times C_2$ formulation. Then we complexify again, to produce a multivariate vector-quaternion group 1, –1, i, $-i$, $i\mathbf{i}$, $-i\mathbf{i}$, $i\mathbf{j}$, $-i\mathbf{j}$, $i\mathbf{k}$, $-i\mathbf{k}$, $\mathbf{i}$, $-\mathbf{i}$, $\mathbf{j}$, $-\mathbf{j}$, $\mathbf{k}$, $-\mathbf{k}$, i, $-i$, j, $-j$, k, $-k$, ii, $-ii$, ij, $-ij$, ik, $-ik$, and 36 real and complex combinations of vectors and quaternions, forming a group of 64, with a related $C_2 \times C_2 \times C_2 \times C_2 \times C_2 \times C_2$ formulation. Because of the reduction of information involved in defining both multivariate vectors and quaternions as cyclic, and in one producing complex, and the other real, products, the $C_2 \times C_2 \times C_2 \times C_2 \times C_2 \times C_2$ formulation can be expressed by the 64 possible combinations of $\pm \mathbf{i}$, $\pm \mathbf{j}$, $\pm \mathbf{k}$, $\pm i$, $\pm j$, $\pm k$, the algebra of the Dirac gamma matrices. Further dualling is possible on the same basis, but it is clear that only three fundamental principles are required to continue the dualling to infinity – opposite signs (or equivalent), the distinction between real and imaginary components, and the introduction of cyclic dimensionality – and to establish every conceivable combination of these, that is to establish every type of dualling, requires a group of 64 elements.

C_2	C_2	± 1	conjugate
C_4	$C_2 \times C_2$	$\pm 1, \pm i$	complexify
Q_8	$C_2 \times C_2 \times C_2$	$\pm 1, \pm i, \pm j, \pm k$	dimensionalize
V_{16}	$C_2 \times C_2 \times C_2 \times C_2$	$\pm 1, \pm i, \pm i, \pm j, \pm k$	complexify
QQ_{32}	$C_2 \times C_2 \times C_2 \times C_2 \times C_2$	$\pm 1, \pm I, \pm J, \pm K, \pm i, \pm j, \pm k$	dimensionalize
VQ_{64}	$C_2 \times C_2 \times C_2 \times C_2 \times C_2 \times C_2$	$\pm 1, \pm i, \pm I, \pm J, \pm K, \pm i, \pm j, \pm k$	complexify

The process becomes entirely repetitive at the level of V_{16}, while VQ_{64} is what we obtain by combining C_2, C_4, Q_8, and V_{16} as independent elements, establishing conjugation, complexification, dimensionalization and repetition. Beyond this stage, we can consider the sequence proceeding through an infinite series of quaternionic structures by repeated processes of complexification and dimensionalization, creating an infinite-dimensional Grassmann algebra, whose units are each quaternionic. Repetition necessarily sets in as soon as we establish the principle of closure, and closure, as we shall see, allows us an immediate procedure for returning to zero. (The process of conjugation, of course, can be repeated, like those of complexification and dimensionalization, but it is defined in such a way that repetition produces no new structure.)

THE PARAMETER GROUP

So far, this sounds purely mathematical. What relevance, then, does it have to physics? The answer is that it is, in fact, purely *physical* in origin. Duality is a *physical* requirement of the description of nature, and not necessarily a requirement of an abstract system of logical thought, though it may well be that such a system cannot be separated from considerations derived from physical requirements. In effect, when we define the dual, we *define the physical.* The words are synonymous. So, we should expect to see manifestations of these structures in physical 'reality', as we ordinarily perceive it.

From purely empirical considerations of physics, it has been possible previously to suggest that it is based on the relationships between only four fundamental parameters: space, time, mass and charge (where charge is a general term for the sources of the electromagnetic, weak and strong interactions) (see Appendix II). Further investigation of these suggest that the most fundamental properties and 'antiproperties' they possess are as follows:

space	nonconserved	real	countable
time	nonconserved	imaginary	noncountable
mass	conserved	real	noncountable
charge	conserved	imaginary	countable

This has the structure of a $C_2 \times C_2$ or D_2 relationship, in which any of space, time, mass or charge may be the group identity element, and each is its own inverse (see Appendix III). It has also been shown that the symmetry is exact, and absolutely unbroken within physics. Especially significant, however, is the fact that countability or discreteness is a necessary requirement for cyclic multidimensionality, for

unidimensionality is an obviously necessary property of a continuous or noncountable quantity – it can't have an origin. But, multidimensionality is also a necessary property of discreteness (at least in a nonconserved parameter like space). Discreteness has to have a reference or origin; we can't imagine observing discreteness in space without at least another dimension for reference. However, when we investigate space and charge, we find further that the dimensionality in each case is also *three*-dimensionality and cyclic, just as we require for our dual system. Space, being real, has the properties of a multivariate vector, with the associated pseudoscalar being imaginary time, in the '4-vector' combination; while charge, being imaginary, has the properties of a quaternion, with the associated real scalar being mass. In the case of the quaternions, also, it is significant that three-dimensionality is the only dimensionality which, *mathematically*, preserves the group structure; the mathematical possibility is determined *at the same time* as the physical. With the arguments already presented, we can additionally say that the origin of the *physical* concepts of continuity and discreteness lie in the duality which requires the creation of a cyclic three-dimensionality in our conceptualization of nature. This is, in fact, what we *mean* by continuity and discreteness.

We can see now that two of the distinctions between the parameters, which we have derived inductively from observed physical characteristics (real / imaginary and noncountable / countable), are identical to the C_2 distinctions which extend the original C_2 duality into complexity and cyclic dimensionality. However, even the original C_2 duality (1 / –1) originated from the act of creating 'something from nothing' (1 from 0), the very definition of *nonconservation*, as is the concept of 'successor' which it implies. So, in principle, our group of space, time, mass and charge has all the elements required to extend physical duality to infinity. And, our choice of the distinction between conservation and nonconservation (in effect, incorporating 0 directly as the never-used 'totality', and leaving –1 as implicitly understood rather than explicit) even allows us to simplify a potential $C_2 \times C_2 \times C_2$ structure into the simpler $C_2 \times C_2$ we have used above, with the added bonus that we can represent it as identity element (or single sign of scalar) plus three 'quaternion' terms, thus creating a powerful mapping of the four parameters onto a quaternion space. Alternatively, we could represent mass and charge as 'conjugated' quantities, in the sense that creation of a + value can only be accomplished at the same time as the creation of an equivalent – value. So the group could be written in a form, in which the property / antiproperty distinctions occur as examples of successive applications of the dualling process (nonconjugated → conjugated; real → complex; nondimensional → dimensional):

space	nonconjugated	real	dimensional
time	nonconjugated	complex	nondimensional
mass	conjugated	real	nondimensional
charge	conjugated	complex	dimensional

Mathematically, it is possible to create a dual set or parameters, one form of which is seen in certain versions of the Dirac theory, in which certain characteristics of space and time, and mass and charge are reversed, for example the real / imaginary characteristics.

space*	nonconserved	imaginary	countable
time*	nonconserved	real	noncountable
mass*	conserved	imaginary	noncountable
charge*	conserved	real	countable

The combined group is then extended to $C_2 \times C_2 \times C_2$, with a quaternion representation (Q_8) with both signs of scalar. This extends the mathematical representational space, but is not needed in the physical representational space, and, in the more sophisticated (quantum field) versions of Dirac theory becomes redundant.

There is also a distinction between the representations of the distinctions between the parameter properties (e.g. real / imaginary) by existence / nonexistence conditions, as here; and the explicit representation of these properties by their explicit natures (e.g. vector / quaternion). The minimum representation in the latter case is of the order $C_2 \times C_2 \times C_2 \times C_2 \times C_2 \times C_2$, or the Dirac group. In the former case, there are at least two striking visual representations of the group relations, which bring out the significance of the C_2 distinctions and of the principle of cyclic dimensionality.

COLOUR REPRESENTATION

The four parameters, space, time, mass and charge, are represented by concentric circles, the parameter chosen as the identity element for the group occupying the centre circle. The division of the properties into three components is reflected by the division of the circles into three sectors. The properties (say, Real, Nonconserved, Discrete) are represented in Figure 1, by primary colours (say, Red, Green, Blue), and the 'antiproperties' (Imaginary, Conserved, Continuous) by the complementary secondary colours (Cyan, Magenta, Yellow). All of these

choices are individually arbitrary (as we see from Appendix III), as is the choice of secondary colours to represent the properties, and primary colours to represent the 'antiproperties' in Figure 2. The division between properties and antiproperties is also a completely free choice. Only the overall pattern is fixed. As configured, with Space selected as the identity element, and the colour representation for the properties selected as indicated, the innermost circle represents Space, the next Charge, the next Mass, and the outermost circle Time. But this will be changed as soon as we redefine any of the colour representations or exchange the status of any or the property-antiproperty pairs. In addition to being an alternative representation of the main group, Figure 2 may also be used, simultaneously with Figure 1, as a representation of the dual group, which can be obtained (for example) by exchanging the status of real and imaginary quantities (as in some versions of the Dirac theory).

Figure 1

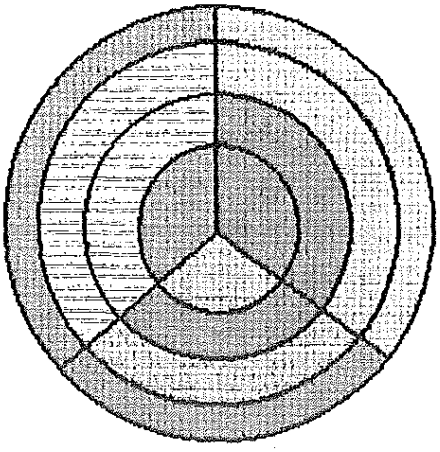
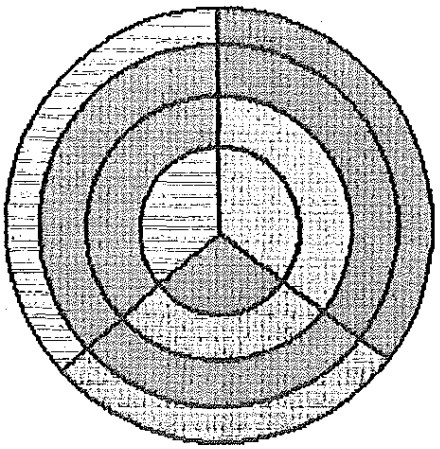


Figure 2



The nature of the fundamental parameter group is demonstrated in this representation by summing up the colour combinations in each of the circles. This results in a white inner circle for the identity element, and a sequence of the three primary colours (Figure 3) or secondary colours (Figure 4), which adds up to a white totality.

Figure 3

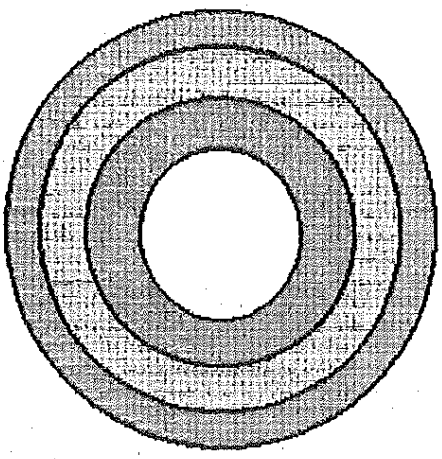
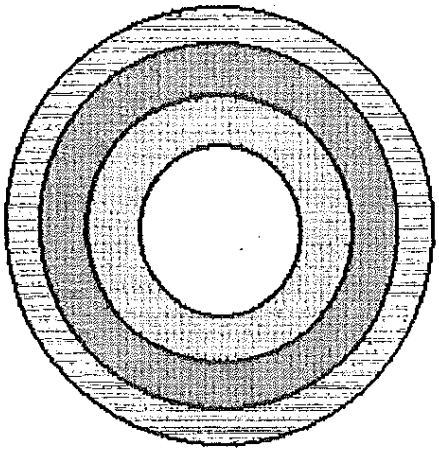
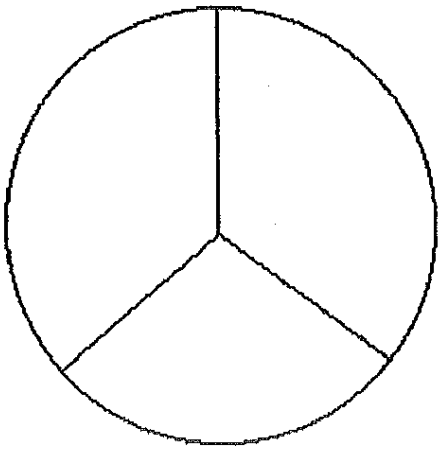


Figure 4



Adding up the property-antiproperty combinations in the sectors also results in a white totality for each sector, as expected (Figure 5).

Figure 5



Clearly, this colour representation derives its effectiveness from the fact that a three-colour system is a simulation of three-dimensionality in an alternative vector or quaternion representation. Such representations are also possible in a more direct form.

3-D (VECTOR) REPRESENTATION

The alternative (vector or quaternion) representation (Figures, 6 and 7) can be seen as a literal interpretation of the x, y and z, used in the tables in Appendix III. The x, y and z directions represent the properties, and the $-x$, $-y$, and $-z$ directions the antiproperties (again, according to an arbitrary choice). The four Red lines, in Figure 6, drawn from the origin

of these 3-dimensional axes, then represent the four parameters, and the Cyan lines those of the dual group. Figure 7 shows the same representation as Figure 6, but without the axes.) The Red lines are reflections of each other in two planes. We can represent these as preserving the sign of the volume element (or identity), if the axes are taken in the same cyclic sense; and so they correspond to the parameter group The Red plus Cyan lines are the reflections of each other in a single plane, and do not preserve the sign of the volume element; and so form the parameter group taken with its dual. The reflection of a line in three planes produces its exact dual.

It will be apparent that the representation of the dualities of the parameter group using either the three real spatial dimensions or the pseudo-dimensions of the three primary colours is a powerful way of bringing out the connections between duality and dimensionality, and the fact that all the individual dualities are, in effect, versions of the same mechanism. It is also a convenient way of showing how the parameter group can be used to represent a kind of 'super-duality' of all the elements, conveniently displayed using the particular duality of dimensionality.

Figure 6

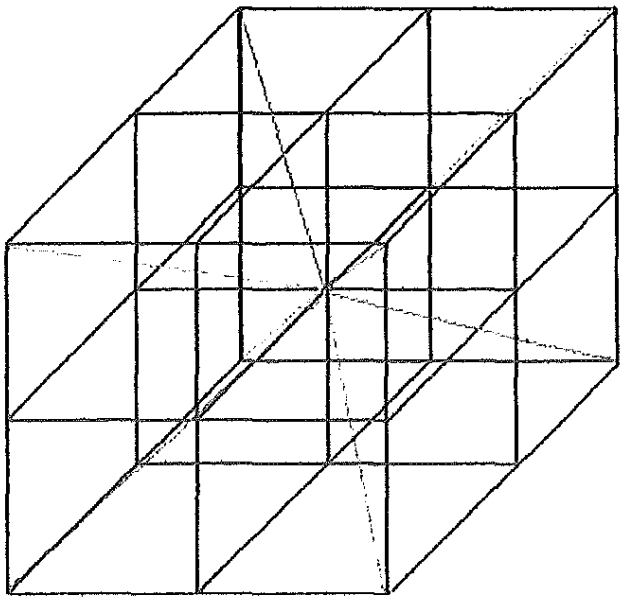
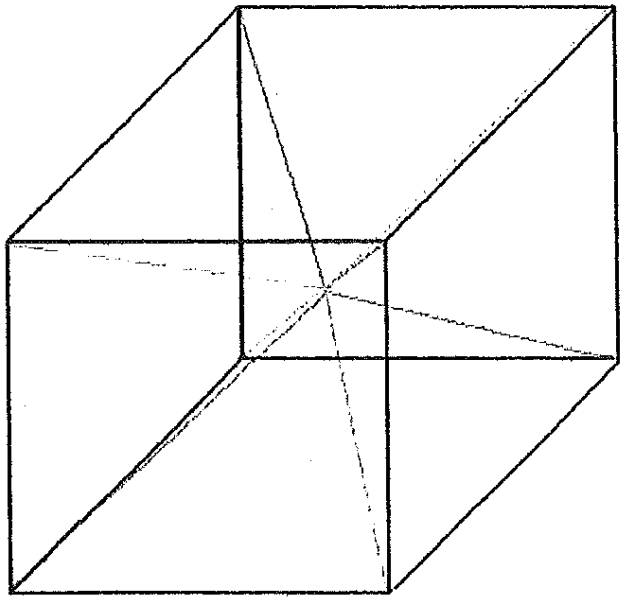


Figure 7

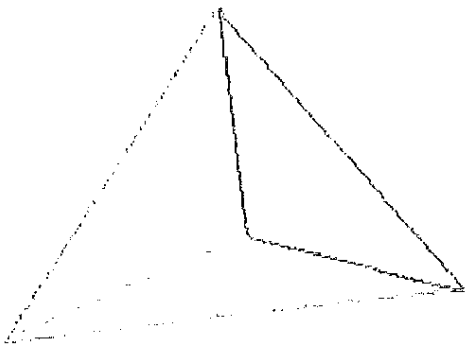


To map between the colour and 3-D representations, we could represent the positive x, y, z axes in Figure 6 by the primary colours, say Red, Green, Blue, against a black background, and the corresponding negative directions by the complementary secondary colours, which, in this case, would be Cyan, Magenta, Yellow. The four Red vectors of Figure 6 would become respectively White, Blue, Red, Green, as determined by the colour coding in Figures 1 and 3, and the four Cyan vectors of Figure 6 would become respectively White, Yellow, Cyan, Magenta, as determined by the colour coding in Figures 2 and 4.

TETRAHEDRAL REPRESENTATION

Yet another 3-D representation (Figure 8) would place the parameters at the vertices of a regular tetrahedron, with the six edges coloured to represent the properties and antiproperties as in Figure 1.

Figure 8



We can consider the faces of the tetrahedron to be the members of the dual group, and, clearly, an alternative representation would reverse primary and secondary colours and / or the roles of faces and vertices. It might be possible to consider the tetrahedron as close-packed with inverted tetrahedra with complementary colour-representation in an all-white solid-space, which can be extended to infinity.

An interesting possibility is that a structure like the one in Figure 8, if flattened out in a 2-dimensional space, could be considered as a 'dart' or 'kite' in a Penrose tiling pattern (the base line being optional to the connections between the vertices representing the parameters). Penrose tiling is, of course, a five-fold symmetry, and, typically, a group of five darts (or kites) will produce a star-shaped pattern with each of the darts joined with all the others at its apex, and with its two nearest neighbours along two of its edges. The star then has five inner and fiver outer vertices, and, surrounding a central star made of darts, we will have ten kites, each joined by two edges to two nearest neighbours, and by another edge to one of the ten outer edges of the star. If we assume that each of the 4 vertices of any dart (3 symmetric and 1 asymmetric) must represent one of space, time, mass and charge, and that joint vertices may only represent one parameter, then putting a 3-dimensional parameter like charge or space, at the centre of the star forces us to choose the inner and outer vertices in such a way that the other 3-dimensional parameter occurs three times in the five inner or outer vertices, while each of the other parameters occurs only once. It is interesting that another five-fold structure (the Dirac algebra) emerging from a combination of two types of 3+1-component units (space-time 4-vectors and mass-charge quaternions) is also forced to 'privilege' one of its two 3-dimensional quantities, space or charge.

THE DIRAC NILPOTENT

In the parameter group, not only are the properties dual, but so is the distribution between the parameters. This is why the minimum representation of the full duality is the Dirac algebra, of order $C_2 \times C_2 \times C_2 \times C_2 \times C_2 \times C_2$, and produced by the 64 possible combinations of the 'double vector', ± **i**, ± **j**, ± **k**, ± ***i***, ± ***j***, ± ***k***. Hidden within this representation, but expressive of the cyclic nature of the operators (and apparent in Figures 6 and 7), are the respective pseudoscalar and scalar terms, ± *i* and ± 1 (and we could, alternatively, use the 'double vector', ± **i**, ± **j**, ± **k**, ± ***ii***, ± ***ij***, ± ***ik***). Leaving out ± signs, the full representations are: space (**i**, **j**, **k**), time (*i*), mass (1), and charge (***i***, ***j***, ***k***). The fact that the Dirac algebra can be derived from a combination of two three-dimensional operators now suggests a further possibility, based on our mathematical and visual representations of the group. This is that one of the two three-dimensional parameters may be mapped on to the other three parameters, represented as the 'dimensions', and, in fact, the smallest set of units from which the full algebra can be derived comes from exactly such a mapping. We take, for example,

time	space	mass	charge
i	**i j k**	1	***i j k***

and, taking, each of the units of charge onto one of the 'dimensions' represented by time, mass and space,

i	**i j k**	1	***i j k***
k	***i***	***j***	

create the following combinations:

ik	***ii ij ik***	***j***

A set of five units of this kind, or pentad, will always generate the entire Dirac algebra of 32 parts (excluding signs). The 32 parts turn out to be 1 and *i*, and six Dirac pentads, three based on the quaternion operators (as here) and three on the vector operators. Any of these sets can be used as the basis for the five gamma matrices in the Dirac equation, but it is most convenient to use the quaternions, as here, because charge is a conserved quantity, and the mathematical structure then has a convenient physical interpretation. (In the case of vector space, the components are not uniquely determined, because the quantity is nonconserved, and can even be arbitrarily reduced to a single one. However, the conservation of

charge is directly related to the conservation of angular momentum, and so brings in the spatial rotation simultaneously, as becomes evident in the full explanation of symmetry-breaking.) The combined units take on the physical characteristics of their component quantities. The charge units introduce conservation and discreteness (quantization) to all the quantities. However, the new conserved quantities retain their respective pseudoscalar, vector and real scalar identities, as Dirac energy, Dirac momentum and Dirac rest mass:

ik	$ii\ ij\ ik$	j
E	$\mathbf{p}$	m

In a nonconserved form they produce the respective quantum operators:

$\partial/\partial t$	∇	m

Treating the momentum term as a single quantity, the free-fermion Dirac state vector now becomes a nilpotent $(\pm ikE \pm i\mathbf{p} + jm)\ e^{-i(Et - \mathbf{p}.\mathbf{r}.)}$, where $(\pm ikE \pm i\mathbf{p} + jm)$ expresses the absolute conservation of charge and mass(-energy), and the exponential term, operated upon by $(\pm ik\partial/\partial t \pm i\nabla + jm)$, the absolute nonconservation of space and time. The identities of the three 'charge' operators are preserved, even in the combinations of $(\pm ikE \pm i\mathbf{p} + jm)$, for they now become discriminated into ones with timelike (weak), spacelike (strong) and masslike (electric) properties, and the effects can be distinguished physically by the aspects of angular momentum conservation to which they relate. The Dirac algebra, which produces the simplest possible combination of all the dualistic properties required by space, time, mass and charge, generates a broken symmetry in the manifestations of the charges' interactions.

Significantly, the term $(\pm ikE \pm i\mathbf{p} + jm)$, which is expressed most conveniently as a row or column vector with four components (yet another 4-vector mapping, and one which can be accomplished with the four quaternion components, $1, i, j, k$, if required) is a nilpotent or square root of zero, a precise expression of the fundamentally dualistic process of returning 'something' back to 'nothing' through a squaring operation. The classic way of doing this is through the Dirac equation in which the nonconservation operator $(\pm ik\partial/\partial t \pm i\nabla + jm)$ is applied to the state vector $(\pm ikE \pm i\mathbf{p} + jm)\ e^{-i(Et - \mathbf{p}.\mathbf{r}.)}$, in which $(\pm ikE \pm i\mathbf{p} + jm)$ represents the conserved terms, so producing a zero result. This expresses the fact that our fundamental duality has been represented in terms of conservation and nonconservation, and that the effect of applying both is to maintain the zero totality.

The Dirac nilpotent represents the most concise packaging of the dualistic information contained in the parameter group, the most complete way of parameterizing nature; and, as we have seen, the combination of all the desired physical elements, with all the inherent symmetries, into a single, parameterization of nature, is the same as the process of 'quantization' of energy, momentum (or angular momentum) and 'rest mass'.

The interaction of a fermion with the (infinite) vacuum, or mass-energy continuum, produces an infinite succession of products or superpositions of Dirac nilpotent states. This extends the dualling processes to infinity. Each of the 'virtual' states produced also acts in the same way, producing a pattern of the same form as the Conway system of constructed real numbers. The requirement of infinite dualling ensures the entanglement of all states in the universe (although, as with classical interference, decoherence will make this virtually unobservable except in special cases).

It seems that we get something from nothing, not just in a physical way, by perfect symmetry between the parameters denying overall characterization, but also literally, by making the fundamental unit of our characterization a square root of zero, and that this becomes zero in the Dirac equation when we apply to it a differential operator, and generate an exact equal to it as an eigenvalue. The Dirac equation itself expresses the fundamental duality of our view of 'nature', for the left-hand term (the differential operator) specifies the nonconserved aspects, and the right-hand term (the wavefunction) the conserved aspects. Any individual nilpotent wavefunction structure ($kE + ii\mathbf{p} + ijm$) must then be unique because a superposition of identical ones would zero the wavefunction of the entire universe, and Pauli exclusion becomes obvious. This specification of uniqueness requires instant correlation, at the same time as the 4-vector nature of the operator requires time-delayed action between discrete sources. It is also a reflection of the uniqueness or local conservation of individual charge components.

THE FACTOR 2

Duality has an astonishingly simple manifestation in physics through the appearance of the factor 2 everywhere where it becomes significant. This is discussed in detail in the paper 'The physical significance of the factor 2' [Rowlands, 2002]. The main result of this is that each process that doubles the options available also produces a doubling of the physical effect which can be reduced to simple numerical terms. At the same time this is often balanced by a halving of the options in another direction.

Thus, when we describe a physical process using constant, rather than changing, terms, we are effectively using both sides of the + / – duality at once. This is the case when we use potential, rather than kinetic energy equations, or both action and reaction sides of Newton's third law, or even relativistic, rather than rest, mass. Relativity itself does not introduce the factor 2, but relativistic equations can often be used as classic examples of changing conditions. The most controversial instance in historical terms is the double bending of light rays in a gravitational field, which can, in fact, be seen as an example of the use of a kinetic, rather than potential, energy equation. Of coursing, halving in one respect may lead to doubling in another. So halving the energy, by using a kinetic term, produces a doubling of the angular deflection; but it is also possible to produce the doubling directly by taking both space- and time-related effects into account. This could be seen as an application of the second route to duality: complexification, or the adding of a complex term to a real one. It is significant here that the group relationship between the physical parameters is so integrated that such apparently alternative explanations emerge without any fundamental contradiction. Both explanations are equally true, and neither has precedence over the other.

Another significant alternative can be seen in the explanation of half-integral fermion spin and its effects. It is possible to derive the resulting magnetic properties using a classical kinetic energy equation, and so using one side of the + / – duality. On the other hand, both the Dirac and Schrödinger equations derive the half-integer value for spin indirectly by using the *doubling* effect produced by the third process of duality: dimensionalization. This comes about in both cases via the anticommuting properties of multivariate momentum vectors, a direct result of 3-dimensionality. What this indicates is that the symmetrical structure applied to physics is organized in such a way that *both these interpretations of the dualling process apply simultaneously.* In effect, this hidden balancing act also operates in yet other, more subtle ways because the virial relation between potential and kinetic energies is specifically one of doubling only when the force laws which apply are those characteristic of 3-dimensional space; and the action and reaction mechanisms which produce the doubled value for potential energy rely on applying vectorial (or dimensional) considerations to the kinetic energy term.

That the doubling mechanism also applies in purely mathematical, as well as in physical, contexts is evident from the topological explanation of the Aharanov-Bohm effect, though the physical and mathematical applications must ultimately have the same origin. Square-rooting and halving have an intimate relationship, which is manifested physically in the relation between vector spin terms of bosons and fermions and their

respective uses of double or single nilpotent operators, in addition to the halving approximation used to find the kinetic energy term in the bionomial expansion for relativistic mass. This relationship is determined entirely by the fact that 3-D Pythagorean addition is a dualistic process, with a numerical doubling arising from noncommutativity, and this applies to both the vector operators used for space and momentum, and the quaternion operators used in the Dirac nilpotent.

There are also other possible mathematical connections. It is tempting, for instance, to believe that the uniqueness of the value ½ as the real part of the zero-solutions of the Riemann zeta function has a significance which is physical as well as mathematical, and that, as Hilbert originally conjectured, the solutions represent the eigenvalues and energy levels of an Hermitian operator, which is the Hamiltonian of a quantum mechanical system. It is conceivable that the ½ is related to the zero-point energy term of a series of fermionic harmonic oscillators. It is certainly true that, solving the Dirac nilpotent equation for any spherically-symmetric potential other than a linear or Coulomb one (i.e. under harmonic oscillator conditions) requires a Coulomb or phase term with numerical coefficient ½, which is of the opposite complexity to the rest of the potential, and which can be associated with the zero-point energy or (equivalently) the random directionality of the fermion spin. There may also be some physical significance in the fact that integers, like the fundamental parameters, only add directly to produce other integers or in the form of squares to produce squares of integers, but do so in an infinite progression. Both of these mathematical results suggest the possibility of further fundamental significance in the factor 2.

QUANTUM PHYSICS AND THE CLASSICAL TRANSITION

The definition of the Dirac nilpotent suggests that this is the most efficient way or parameterizing nature while ensuring its total 'nothingness'. It may be possible to relate this to the aims of *topos* theory, in using a nilpotent Pythagorean structure to create a 'parameter space' which contains within itself dynamical and other physical possibilities. The uniqueness of the individual Dirac nilpotents, together with their necessary entanglement with each other and their infinite interaction with the vacuum, suggest that this is a real number space, with the numbers countable in the Robinson or Löwenheim-Skolem sense (see Appendix III). Through the Dirac equation, the nilpotents are then interpreted simultaneously in terms of conservation (the eigenvalue) and nonconservation (the operator). Quantum physics thus becomes a natural consequence of the fundamental meaning of conservation and nonconservation, and its separation from the physics of measurement

(classical physics) becomes obvious. It is still necessary, however, to make sense of the classical transition, and also of the relationship between gravity and the other forces. Considerations of such ideas may also suggest the origin of the classical laws of thermodynamics.

Measurement processes are discrete, and involve discrete sources (or charges). They rely on the $SU(3) \times SU(2) \times U(1)$ symmetries which apply to these sources (and whose direct expressions are 'interactions', equivalent in principle to the action of classical field terms) producing restrictions on the freedom of the individual wavefunctions to contain infinitely possible variations in space and time coordinates.

A hypothetically isolated system (e.g. a hydrogen atom not interacting with other hydrogen atoms) must be purely quantum. Once we have any classical element or interaction the system is no longer isolated. This is how we make a measurement. We can't make classical-type observations on an isolated system, otherwise it wouldn't be isolated. An isolated system conserves E-**p**-m within the system, linking it with the total k, i, j charge values, whether 0 or unit, positive or negative. This system must remain coherent – with angular momentum operators aligned, so that addition is effectively scalar, like that of the charge units. If the system interacts with an external system, then it can no longer be defined in an isolated way: the connection between the conservation laws for charge (k, i, j) and angular momentum (E-**p**-m) is broken. If the system is not isolated, then energy is not conserved *within* it, but some is lost to the 'rest of the universe' with which it interacts. Hence, we need the second law of thermodynamics, and, in fact, the first law (where the energy balance is only maintained globally by incorporating 'lost' energy into the equation). The connection between the second law of thermodynamics and the direction of time is now apparent. To make a measurement requires a semi-classical situation with a non-isolated system; as soon as we make a measurement, we lose energy from the system to the 'rest of the universe', so increasing the 'entropy'. The sequence of events behaves as an irreversible sequence because time itself is irreversible, because of its continuity, and a sequence of event 'measurements' must follow the same sequence; but this for any known pair of events will always require an increase in entropy.

CONCLUSION

Physics, mathematics and philosophy emerge together out of the basic idea of duality, though the concept applied is more fundamental than this name, with its connotations of a necessary discreteness, would imply. Mathematics is not something 'applied' to physics for 'convenience'. It is, in fact, extremely *inconvenient*, as the mathematical laws of physics

are general differential equations, which have to be reinterpreted ('solved', using different boundary conditions) every time a measurement is taken. Observation and theory, in physics, necessarily use incompatible types of mathematics because observation depends only on one member of the parameter group (space), while theory sets up the properties of the other members in opposition. So, mathematics is required only because it is a fundamental component of physics, and the structure of mathematics itself seems to suggest physical boundaries to the type of ideas which can be made mathematically useful (though obviously not in the form of a purely one-to-one correspondence). In fact, physics can become a kind of test of the ultimate value of mathematical structures at the fundamental level.

For example, physics appears to insist on the fact that all discrete quantities must be dimensional. This would not be required of a mathematical theory based, as most are, on the primacy of the integer series. However, if we begin mathematics with the integer series, then we have major problems in accommodating the reals – there is no natural progression – and, if we assume, like most axiomatists, that the most fundamental proposition in mathematics is $1 + 1 = 2$, we will come up against the problems that Gödel identified with axiomatic theories or 'rigidly logical systems' which are intrinsically incomplete.

However, in physical terms, we may suppose that the integer series is not primary and that arithmetic, although the most psychologically familiar, is not the most fundamental branch of mathematics; and, further, that, the moment we assume that the number 1 (or even number at all) is the most basic concept in mathematics (or indeed in human thought), we have at the same time brought in a whole package of information that we will never be able to establish from first principles. Physics, in fact, tells us that integers and discrete numbering are not primary – they are associated with dimensionality – and dimensionality only has a meaning in the context of complexity. The integers are really a *codification* of a multiplicity of prior stages in mathematical evolution. To begin with them will necessarily produce an incompleteness in our logical procedures, with key steps appearing merely as assumptions in a circular argument. But, if we begin at the true primary stage, with a zero end product at every stage, we effectively remove the incompleteness in our axiomatization. We also reach a primary stage in which even the word 'dual' loses its meaning, although its convenience for the later stages makes it worth retaining if separated from its numerical associations.

The very applicability of the concept of 'duality' to the process of returning from 'something' to 'nothing' implies that the actual processes of counting and generating numbers are created, along, with 'addition', 'squaring', and other arithmetical procedures, at the same time as the

categories of conjugation, complexification, and dimensionalization are separated from their dualistic counterparts. Defining the integers as an ordinal set within a much more fundamental process allows us to create new mathematical processes in which this ordinal set is applied in other ways, and so we can create types of mathematics where the relation to physical categories is less direct, but the ultimate 'physical' or 'dual' origin will remain.

From a purely physical point of view, the Dirac nilpotent would appear to be the perfect way of producing something from nothing; its structure also effectively incorporates or generates all the discrete and continuous groups of interest in fundamental physics, from C_2 to E_8 [Rowlands, Cullerne and Koberlein, 2001]; while the infinite imaging of the fermion state in the vacuum and the infinite entanglement of all nilpotent fermion states extends the dualling to infinity, as required. At the other end of the scale, the author, and collaborators, have shown, in many previous papers, how this concept applies to the structure of fundamental particles and the four fundamental physical interactions.

Appendix I Quaternions and multivariate vectors

Quaternions follow the multiplication rules:

$$\begin{gathered} i^2 = j^2 = k^2 = -1 \\ ij = -ji = k \\ jk = -kj = i \\ ki = -ik = j \\ ijk = -1 . \end{gathered}$$

If the quaternions are complexified we have:

$$\begin{gathered} (ii)^2 = (ij)^2 = (ik)^2 = 1 \\ (ii)(ij) = -(ij)(ii) = i(ik) \\ (ij)(ik) = -(ik)(ij) = i(ii) \\ (ik)(ii) = -(ii)(ik) = i(ij) \\ (ii)(ij)(ik) = i . \end{gathered}$$

Multivariate vectors follow exactly the same multiplication rules:

$$\begin{gathered} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = 1 \\ \mathbf{ij} = -\mathbf{ji} = i\mathbf{k} \\ \mathbf{jk} = -\mathbf{kj} = i\mathbf{i} \\ \mathbf{ki} = -\mathbf{ik} = i\mathbf{j} \\ \mathbf{ijk} = i . \end{gathered}$$

In effect, this means defining a 'full product' for two vectors **a** and **b** of the form

$$\mathbf{ab} = \mathbf{a}.\mathbf{b} + i\,\mathbf{a} \times \mathbf{b}\,.$$

The rules for multivariate unit vectors are also exactly identical to those for Pauli matrices, and, through the additional cross term, immediately generate the concept of fermion spin.

Appendix II: Properties and antiproperties of space, time, mass and charge

The inextricable combination of properties and antiproperties means that neither an epistemological conception of 'reality' (we create it by our perception) nor an ontological one (it is 'out there' waiting for us to discover it) is meaningful. The division between epistemology and ontology can be based on the opposition of parameters susceptible or not susceptible to measurement (space versus the rest), or, alternatively, on the opposition nonconserved and conserved parameters (space and time versus mass and charge). In either case, the complete description requires both – 'nothing' is neither epistemological nor ontological. The properties and antiproperties also incorporate all the fundamental types of 'mathematical' number: positive, negative, integer, rational, algebraic, complex, transcendental, denumerable real, nondenumerable real, fixed and variable. They also include a concept of absolute 'uniqueness', which has not yet found its way into conventional mathematics, unless in the properties of the Dirac nilpotent algebra.

(a) Nonconserved / conserved

Physics structures itself by defining systems in which conserved quantities remain fixed while nonconserved quantities vary absolutely. Differential equations show the variation of space and time coordinates while retaining the fixed values of mass and charge, and the quantities which depend upon them: energy. momentum and angular momentum. Both nonconservation and conservation are absolute. The Dirac equation for a free fermion expresses this fact in its most convenient form. Quantum mechanics, in this form, is more explicable than classical mechanics, in that it fully expresses the nonconservation properties of space and time.

Conservation applies to all three types of charge. Lepton and baryon conservation are obvious manifestations, respectively, of pure weak and strong charge conservation, as is the non-decay of the proton.

Conservation laws of mass and charge is also *local*, rather than global. Classically, each element of mass or charge as having a permanent *identity*. Nonconservation is exactly opposite: space and time elements have no identity whatsoever. Hence, space and time have translation symmetry, with their elements specifically *stated to be indistinguishable* in physical equations. Three-dimensional space also has rotation symmetry; that is no identity for spatial *directions* or unique set of dimensions. The contrasting properties of mass and charge are 'translation asymmetry' (conservation of quantity), and 'rotation asymmetry' (electromagnetic, weak and strong charges independently conserved). Charge, unlike space, is conserved in both quantity and 'direction' (i.e. type). Weak, strong and electric charges are not interconvertible. The axes are fixed, along with the units.

The translation and rotation symmetries, of course, are identified by Noether's theorem with conserved quantities. Time and space translation symmetry are identified respectively with energy (E) and momentum ($\mathbf{p}$) conservation, while space rotation symmetry becomes identical to the conservation of angular momentum ($\mathbf{J}$). These three conservation laws can be identified further with the conservation laws of mass, value of charge and type of charge, and, in fact, the additional conserved quantities (E, $\mathbf{p}$, $\mathbf{J}$) can be seen as being 'created' at the same time as the application of the quaternion operators associated with the conserved w, s, e charges to the parameters time, space and mass produces the Dirac state. The conservation of charge type (w, s, e), alternatively rotation asymmetry or charge independence, manifests itself in the mutual independence of the three different aspects of angular momentum conservation (handedness, direction, and magnitude).

Previous work by the author and colleagues has shown that fundamental particles may be defined in terms of their w, s, e charges and rest mass, with the last determined ultimately from the charge structure. The conventional definition of a fundamental particle assumes an irreducible representation of the Poincaré group, or the group of space and time translations and rotations compatible with special relativistic invariance. Here, it can be seen that such translations and rotations are essentially identical to the conservation properties related to charge and rest mass which define a particle in the present theory.

Gauge invariance is a further demonstration of the absolute nonconservation of space and time, and, according to the Yang-Mills principle, is as local as all the principles of conservation. A system which is conservative in relation to mass, charge, energy, momentum or angular momentum, will remain so under arbitrary changes of the coordinates representing the nonconserved quantities, space and time.

(b) Real / imaginary

Pythagorean addition, or addition through squared values, is important to all the fundamental physical quantities. This is a consequence of their origins in quaternion and 4-vector representations. 4-vectors, with three real parts and one imaginary, are a familiar representation of Minkowski space-time. If the vector (or real) part is multivariate, then spin is automatically included. The three components of charge (say, *ie*, *js*, *kw*) can be considered as the 'dimensions' of a single charge parameter, with their squared values used in the calculation of forces added, in the same way as the three parts of space, by Pythagorean addition:

space-time	**i***x*	**j***y*	**k***z*	*it*
mass-charge	*ie*	*js*	*kw*	*m*

The opposing real and imaginary natures of space and time explain simply why identical masses attract, while identical charges (of any kind) repel, the coupling strength producing respective positive and negative values.

Real numbers can be privileged according to sign; imaginary ones cannot, and there are always simultaneous and equal status + and – solutions to consider. Thus, while real mass can be made unipolar (to ensure that it remains a continuum), imaginary charge always produces solutions of two signs, and 'antifermions' and 'antibosons' (with opposite signs of electromagnetic, strong and weak charge) have the same status as 'fermions' and 'bosons'. For the same reason, imaginary time has two possible signs, of equal status, in physical equations, though, as a continuous quantity, it has only one physical direction, and cannot be reversed.

Charges, as imaginary quantities, are only accessible through their squared values in interactions; mass, as a real quantity, is accessible in terms of its unsquared value (as inertia) as well as through its squared value (as gravitation). Time, as an imaginary quantity, is only accessible physically through its squared value (in acceleration), while space, like mass, is accessible through both its squared and unsquared values. Time 'measurement' always requires acceleration (because uniform velocity is imaginary), while space measurement is always direct. Time also is always the independent variable in physical equations, because we have no direct control over it, while space is the dependent variable.

(c) Countable / noncountable

Noncountable or continuous time and mass-energy have no origin. Continuous energy (vacuum) requires an infinite universe. Continuous time requires one without beginning or end. Continuous quantities cannot be reversed, because they have no origin. Mass becomes unipolar (with a single sign) while time is unidirectional. However, time as an imaginary quantities, has two *mathematical* solutions in equations. Hence, there are two directions of time *symmetry*, while there is only one direction of physical time. The unipolarity of mass is the reason why we have a CPT, rather than an MCPT, theorem, with C standing for charge conjugation, P for space reflection and T for time reversal, each of which has two mathematical sign options.

Only discrete quantities can be multidimensional, because multidimensionality requires origins, even if they aren't fixed (as in the case of space). Also, the discreteness of a quantity like space, with unfixed origins, is only possible through dimensionality. What we call 'measurement' in space requires discontinuity in both quantity and direction, and includes both reversals and changes in orientation. This is why what we call 'measurement' takes place only through space. 'Time'-measuring devices all rely on some kind of repetition of a spatial interval. Special conditions, relying on spatial reversibility and dimensionality, have to be used to set up such measurements, although anything which can be perceived at all can be used to measure space, at any time.

There are two definitions of real numbers in mathematics. In Robinson's non-standard analysis, Skolem's non-standard arithmetic, and non-Archimedean geometry, the reals are denumerable. The Löwenheim-Skolem theorem is significant here, in requiring any consistent finite, formal theory to have a denumerable model, with the elements of its domain in a one-to-one correspondence with the positive integers. In the Cantorian definition, which is related to the standard versions of analysis and geometry, they are a non-denumerable continuum. Essentially, the first definition reflects the properties of space, while the second accords with the properties of time (or mass). Both systems yield identical results, because differentiation is a property linked to nonconservation, and not concerned, in principle, with the difference between absolute continuity and indefinite divisibility. Zeno's paradoxes are solved only by assuming a divisible (or digital) space and a non-divisible (or analogue) time. The solution by limits can only be used if we adopt the time-like, or standard, version of differentiability.

Mathematicians have been unable to decide whether mathematics is ultimately continuous or divisible. This is because both are physical options. The indefinite divisibility of space, though sometimes wrongly

termed 'continuity', is a different attribute from the absolute continuity of time. Space's indefinite elasticity, its 'continual' recountability, or unending divisibility, are ways of expressing its *nonconserved* nature, the nonfixed nature of its units, but infinite divisibility is, mathematically, the very antithesis of absolute continuity, and the whole process of measurement would be impossible if space, unlike time, were not divisible in this way. Continuity, again, is not related to differentiability – the confusion here arises from the fact that one of the differentiable quantities, time, *is* continuous, and that the continuity of time is a significant contributor to our psychological perception of states of change. Differentiability is, again, a manifestation of nonconservation, a wholly separate physical category, and is equally valid in both discrete and continuous, and in classical and quantum contexts, though it is obviously not valid where the discreteness is fixed (which is where much of our psychologically-based notion of discreteness comes from).

Wave-particle duality and the opposing Schrödinger and Heisenberg versions of quantum mechanics are the result of adopting predominantly continuous or discrete options for physical quantities. Neither is actually physically possible, and each incorporates the alternative in the process of 'measurement'. The non-quantum-field version of the Dirac theory partially overcomes this by including terms equivalent to those in the dual group, with aspects of the real and imaginary nature of space, time, mass and charge reversed. The nilpotent or quantum field version of the Dirac theory, however, incorporates both discrete and continuous options as duals within its mathematical structure, and no longer requires explicit use of the dual group.

Hamilton was correct, in 1843, in seeing quaternions as being responsible for the three-dimensionality of space, as quaternions introduce the concept of three-dimensionality which vectors subsequently adopt. They also occur one stage earlier than vectors in the evolution of physical concepts via duality. It is significant that the quantized parameters (E, $\mathbb{p}$, m, and, collectively, $\mathbb{J}$) emerge from the application of the charge quaternions to the originally non-quantized space, time and mass. Although quantization may thus appear to be equivalent to converting these three parameters into discrete forms, the discrete versions should be seen rather as *composites*, with which the discrete charge structures are inextricably linked.

(d) The construction of physical laws

The parameter group leads naturally to laws of physics, based on the explicit specification of what is conserved and what is nonconserved. The construction of these laws relies on the fact that every statement about

conservation is simultaneously a statement about nonconservation. To relate the conserved and nonconserved aspects, we use the scaling relations between the parameters, and their dual inverses, which are established at the moment that we compress the eight units of the independent parameters into the five of the combined Dirac algebra, and create composite parameters with characteristic aspects of each. The most significant of these are the ones which relate mass and charge, respectively, to time and space, and which can also be shown to be conserved parameters: namely, energy, linear momentum and angular momentum. These are quantities of the same mathematical form as time and space, which are conserved in exactly the same way as those quantities are nonconserved, and form conjugate pairs with them for exchanging statements about conservation into statements about nonconservation and vice versa. Both classical and quantum laws can now be constructed in terms of these conjugate pairs (e.g. via Poisson brackets).

Energy is a pseudoscalar and is conserved in quantity and individual element (i.e. is translation asymmetric) in precisely the same way as the pseudoscalar parameter time is not conserved in quantity and individual element (i.e. is translation symmetric). It may be regarded as the link between time and the real scalar quantity mass (the gravitational source), as the conservation of energy is directly linked to the conservation of mass. Linear and angular momentum are, respectively, vector and pseudovector, and are conserved in quantity and individual element (i.e. are translation and rotation asymmetric) in precisely the same way that the vector parameter space is not conserved in quantity and individual element (i.e. is translation and rotation symmetric). They may be regarded as the respective links between space and the quantitative values and the quaternion operators applied to the different charge types.

Energy and momentum are regarded as a pure 4-vector in special relativity, but, in a quantum system, this is not strictly true, because quantum (Dirac) energy and momentum terms are only fully represented mathematically when each has a (different) quaternion operator applied to its respective pseudoscalar or vector. When taking the invariant scalar product, of course, these operators disappear, but their significance becomes apparent when we introduce a third term, with yet another quaternion operator (rest mass) to convert them into a nilpotent. A pure 4-vector could not be made into a nilpotent in this way. If we take the Dirac differential operator ($\partial^{\prime}$), we can see that this, also, is not a pure 4-vector, and the same must apply, in general, to relativistic time and space, whether the system is quantum or classical. The nilpotent is completed with a third term, which occupies the same position as rest mass does in completing the differential operator nilpotent. This term is a real scalar,

like rest mass, and its squared value must exactly cancel the scalar product of the time and space components. The fact that it can never be negative means that only retarded solutions are possible for the space-time combination. It also means that, numerically, this new term (the 'proper time', τ) is equivalent to the time value with a zero space component. 'Proper time', however, is, strictly, a rest mass-related, rather than time-related concept. Its validity in both classical and quantum contexts is an indication that the link between these two domains is essentially through the scalar additive nature of the rest mass or 'inertia' of the component systems.

Appendix III: Algebraic representations of the parameter group

A simple algebraic representation of the parameter group can be accomplished by representing the properties of space (real, nonconserved, divisible) by, say, x, y, z, with the respective antiproperties (imaginary, conserved, indivisible) represented by $-x$, $-y$, $-z$. The group now becomes:

space	x	y	z
time	$-x$	y	$-z$
mass	x	$-y$	$-z$
charge	$-x$	$-y$	z

With group multiplication rules of the form:

$$x * x = -x * -x = x$$
$$x * -x = -x * x = -x$$
$$x * y = x * -y = 0$$

and similarly for y and z, we can establish the standard D_2 group multiplication table with space as identity element, and each element its own inverse (the duality of space-time elements and their inverses is, interestingly, a feature of string theory). However, we could just as easily have chosen mass as the identity element by representing the properties and antiproperties:

space	x	$-y$	$-z$
time	$-x$	$-y$	z
mass	x	y	z
charge	$-x$	y	$-z$

and the same applies to time and charge. The 'multiplication' rule here only concerns the signs, but the creation of a Dirac nilpotent incorporating both the space-time 4-vector and the mass-charge quaternion suggests that there must also be a direct multiplication rule between the units of the parameters and those of their inverses, which is exactly what we need to require the existence of the fundamental constants G, h and c.

REFERENCES

Atkins, P. [1994], *Creation Revisited*, Harmondsworth, Penguin, p. 23.

Rowlands, P. [1983], The fundamental parameters of physics. *Speculat. Sci. Tech.*, 6, 69-80, 1983.

Rowlands, P. [1999], *The Fundamental Parameters of Physics: An Approach Towards a Unified Theory*, PD Publications, Liverpool.

Rowlands, P. [1999], Physics; let's get down to basics, in K. Bowden (ed.), *Aspects II*, 1999 (*Proceedings of XX ANPA Conference*, Cambridge, September 1998), 123-134.

Rowlands, P. [2001], A foundational approach to physics, arXiv:physics/0106054.

Rowlands, P. [2002], 'The physical significance of the factor 2', ANPA Philosophical Papers, presented August 2001.

Rowlands, P. and Diaz, B. [2002], A universal alphabet and rewrite system, arXiv:cs.OH/0209026.

Rowlands, P., Cullerne, J. P., and Koberlein, B. D. [2001], 'The group structure basis of a foundational approach to physics', arXiv:physics/0110091.

Young, N. [1988], *An Introduction to Hilbert Space*, Cambridge University Press, p. 59.

Quantum Geometric Algebra

Version 1.1 Jan 2003

Douglas J. Matzke
Lawrence Technologies, LLC
Dallas, Texas 75254, USA
matzke@ieee.org

Michael Manthey
P.O. Box 846
Crestone, Colorado 81131, USA
manthey@acm.org

C. D. Cantrell
University of Texas at Dallas
P.O. Box 830688, Mail Stop EC33
Richardson, Texas 75083-0688, USA
cantrell@utdallas.edu

Abstract

Quantum computing concepts are described using geometric algebra, without using complex numbers or matrices. This novel approach was developed in the first author's Ph.D. dissertation in Electrical Engineering at University of Texas at Dallas (May 2002). This research was built upon the mathematical and conceptual foundation of co-occurrence and co-exclusion previously developed by the second author. Using a topologically derived algebraic notation that relies only on addition and the anticommutative geometric product, a qubit is simply a co-occurrence of two orthonormal state vectors. With this qubit definition, this paper describes the following quantum computing concepts: bits, vectors, states, orthogonality, qubits, classical states, superposition states, spinor, reversibility, unitary operator, singular, entanglement, ebits, separability, information erasure, destructive interference and measurement. These central quantum concepts can be described simply in geometric algebra, thereby facilitating the understanding of quantum computing concepts by non-physicists and non-mathematicians. In fact, after exhaustively analyzing all the discrete qubit states using the geometric algebra notation, it appears there is no other meaning for a co-occurrence of two state vectors other than as a qubit.

1. Introduction

Quantum computing has received significant attention since the announcement of Shor's algorithm [1], which demonstrates that quantum computers can solve some extremely computationally intensive problems more efficiently than any classical algorithm. Unfortunately, hardware and software engineering for quantum computers requires different sets of skills from either research on the physics of quantum computing or hardware/software engineering for traditional computers. The goal of this paper is to lay a foundation for hardware/software design for quantum a computer that is accessible to traditional engineers and computer scientists.

For newcomers to quantum computing the learning curve is steep for two primary reasons. First, quantum computing is based on the principles of quantum physics and is typically expressed mathematically using complex Hilbert space, which is a high-dimensional, complete, vector space, using complex numbers and matrices. The matrix notation is concise and compact, but also opaque to non-mathematicians. Second, quantum computation has many new information concepts that do not naturally arise in classical computing and are therefore unintuitive to traditionally trained engineers and programmers. The difficulty of understanding these new concepts is compounded by the use of Dirac's "bra-ket" notation [2], since the reader must first comprehend the foreign-looking mathematical notation. This article takes the approach of focusing on quantum computing concepts while relying on the notationally simpler geometric algebra [3,4], which uses neither explicit complex numbers nor matrices.

This article is targeted at engineers and programmers with a basic understanding of computer science and mathematics who are interested in learning about quantum computing. From this perspective, quantum computing is nothing more than an information system with very particular "bit" properties and the approach of this relatively short article is to show the design of a mathematically oriented process structure that naturally represents and models these properties. The key bit and quantum-properties and their *relatively simple mathematical representation using geometrical algebra* will be introduced when required and only as needed. Bits form the building blocks of the computing industry and computer professionals have very strong intuitions about them, so this article begins with that perspective.

2. Bits Represented as Vectors

A *bit* expresses a binary *distinction*, the smallest unit of information, and is physically the space reserved (or bit capacity in a disk, memory, register, or communications channel) for a single *binary state* value. The typical choice is to use the implementation-specific values 0/1 to symbolically represent mutually exclusive state pairs such as False/True, dark/light, or male/female. Each binary-valued bit is usually given a symbolic name such as a or b to facilitate describing how multiple bit-states causally interact (i.e. c = NOT a, d = a AND b, using the Boolean algebra conventions with the standard logical operators NOT, AND, OR, etc). A *classical bit* can only have two possible *complementary* states and most importantly, these states are required to be *mutually exclusive*. For example, if states a = True and NOT a = False, then bit a cannot simultaneously be both True and False. Multiple bits can be *concatenated* to express N = 2^n unique states, where n is typically 8 (a byte), 16 or 32 bits. The N resulting states are also mutually exclusive.

The defining properties of classical bits (i.e. a, b, c) are: 1) as above, the complementary state pair is mutually exclusive and 2) the state of each bit can be independently changed. These precise properties can be mathematically represented using vectors (i.e. $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ - in bold font), where each bit is denoted by a distinct vector. This simple choice of representing bits as vectors has many formal mathematical consequences that will be described in footnotes so as not to

disrupt the flow of the article. Two *orthonormal* vectors (orthogonal[i] and unit length) are graphically displayed in Figure 1 as a horizontal and a vertical line that define a plane.

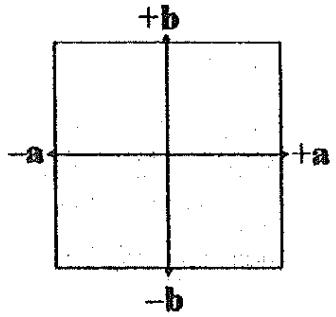


Figure 1. Two orthonormal vectors **a** and **b**

A binary *state* is represented here using the ±*orientation* or *direction* of the bit *vector* and *not* its length, which is always *one*. Using vector **a**, bit state a = True can therefore be denoted as $+\mathbf{a}$ (orientation +1) and NOT a = False can be denoted as $-\mathbf{a} = \overline{\mathbf{a}}$ (orientation −1). The scalar orientation coefficient c preceding the vector $c\mathbf{a}$ can have the *real values* of $c = +1, -1$, or 0, which naturally leads to a *ternary* state system (similar to tristate logic) with *symmetric binary states* of +1 = + and −1 = −, whereas $0\mathbf{a} = 0$ indicates vector **a** has no presence. The choice of mapping bits/states into vectors/orientations defines a binary representation that is a *formal linear system* and can be shown to be Boolean complete [4]. A linear representation is important when building such a bridge between computer science and physics [5].

Table 1 shows how to define the traditional "addition" operator, denoted as +, for this linear algebraic system. The addition of two vectors can be visualized as the address of a point in the plane of Figure 1, but the vector orientation coefficients follow the usual scalar addition rules in Table 1. This algebra is limited to the set of unit scalar values {0,+1,−1}, because this limited scalar set is sufficient to express all necessary distinctions. Table 1 represents modulo 3 addition because repeatedly adding +1 produces the sequence of values 0 => +1 => −1 => 0. This choice is isomorphic to the modulo 3 set of values {0, 1, 2} but is symmetric around 0. Addition of any elements in the algebra always produces another element in the algebra.

Table 1. Scalar Addition table for a + b = b + a

a + b	b = 0	b = +1	b = −1
a = 0	0	+1	−1
a = +1	+1	−1	0
a = −1	−1	0	+1

Using Table 1, an important property[ii] for addition is $\mathbf{a} + \mathbf{a} = -\mathbf{a} = \overline{\mathbf{a}}$. This codifies the existence of an additive inverse, *complement*, or *negation* state for each state in the algebra. Mutual exclusion of two complementary states can now be expressed as $\mathbf{a} + \overline{\mathbf{a}} = 0$, which means the vector **a** can *only point in one direction* at a time because the value 0 *has the special meaning of*

---

[i] With • = inner product; $\mathbf{a} \bullet \mathbf{b} = 0$ means **a** and **b** are orthogonal and $\mathbf{a} \bullet \mathbf{a} = 1$ means collinear

[ii] With mod 3 arithmetic then $2\mathbf{a} = -\mathbf{a}$ because $2\mathbf{a} = \mathbf{a} + \mathbf{a} = -\mathbf{a}$ so $2 = -1$ and also $\mathbf{a}/2 = \mathbf{a}/-1 = -\mathbf{a}$

cannot occur. The symmetric +/– states naturally describe the *destructive interference* of bit vectors, which is critical for quantum computing. Ternary logic is different from traditional Boolean algebra (states of 0/1) because the latter has *no third state*, and hence confounds the states "the opposite of one" and "nothing". Since addition is commutative ($\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$) but subtraction is not ($\mathbf{a} - \mathbf{b} \neq \mathbf{b} - \mathbf{a}$), we use the convention that the sign of the coefficient must always be associated with the particular element, for example, $\mathbf{a} - \mathbf{b} = \mathbf{a} + \overline{\mathbf{b}} = \overline{\mathbf{b}} + \mathbf{a} = -\mathbf{b} + \mathbf{a}$.

The interpretation of 0 to mean *cannot occur* [6] is subtle, yet *conceptually meaningful*, and has these consequences. First, since 0 means *cannot occur* then a scalar multiplication of state vector by zero, such as $0\mathbf{a} = 0$, simply means that the vector $\mathbf{a}$ does not occur or exist and can be removed without any additive effect on an expression (i.e. $\mathbf{x} + 0\mathbf{a} = \mathbf{x}$). Therefore, the *highlighted* cells in Table 1 focus on the addition rules to *input* states with only the non-zero binary values. These cells can be summarized as the rule: *like states invert and differing states pair-wise cancel.* Second, assigning a state equation to 0 and then solving for the roots determines the orientation values of states that *cannot occur* thereby representing the *non-solutions* of the system, which is the opposite of the conventional meaning. Third, in order for two vectors to exactly cancel, they must be *simultaneous*, so *addition means concurrency in time*. This interpretation is consistent with the non-causal nature of quantum computing states. Addition of states is called a *co-occurrence* [6] because it is impossible to distinguish between (or count) two *identical tokens* unless they are presented exactly concurrently. Two such identical tokens presented together represent 1 bit of information because it is impossible to know how many truly exist when presented sequentially [6]; *{q.v.}* chapter 4.

With this brief groundwork in place for classical bits, mutually exclusive states, state inversion and the addition operator (with its interpretation of *concurrency*), the following section introduces how to represent a *qubit* (or quantum bit) plus the other properties required to change the qubit state.

3. A Qubit Represented as the Sum of Two Vectors

A *qubit* requires four states rather than the two states represented by a classical bit, yet still represents only one classical bit because the vectors are constrained to be *redundantly encoded.* Therefore, a minimum of two classical bit vectors $\{\mathbf{a0}, \mathbf{a1}\}$ must be used to represent those four possible states. Since the two distinct and orthonormal bit vectors must both *simultaneously* be allowed to have any binary value, the obvious proposal for a qubit uses addition with all possible non-zero vector orientations:

$$\text{Qubit: } A = \pm\mathbf{a0} \pm\mathbf{a1} \qquad (1)$$

There are four possible variations of signs for this sum and they are assigned the state labels $A_0 = +\mathbf{a0} - \mathbf{a1}$, $A_1 = -\mathbf{a0} + \mathbf{a1}$, $A_- = -\mathbf{a0} - \mathbf{a1}$, and $A_+ = +\mathbf{a0} + \mathbf{a1}$, whose meaning will soon be obvious. Similar to the process used for a single vector, we can show that $A_0 + A_1 = 0$ or $A_0 = -A_1$ because $(\mathbf{a0} - \mathbf{a1}) + (-\mathbf{a0} + \mathbf{a1}) = 0$, which means that states A_0 and A_1 are mutually exclusive. Likewise, states A_+ and A_- are mutually exclusive, because $A_- + A_+ = 0$ or $A_- = -A_+$. As with A_0 and A_1, the states A_+ and A_- are pair-wise collinear with the origin (and later these two sets themselves are shown to be orthogonal). Table 2 follows directly from Table 1 when all possible combinations

of non-zero orientation values are analyzed (only highlighted cells in Table 1). It is convenient to think informally of this table as the non-solutions from solving $A_x = 0$, where the vector coefficients destructively cancel.

Table 2. Valid qubit states highlighted for ±a0 ±a1

Row$_k$	a0	a1	$A_1 = \overline{a0}+a1$	$A_0 = a0+\overline{a1}$	$A_+ = a0+a1$	$A_- = \overline{a0}+\overline{a1}$
R_0	–	–	0	0	+	–
R_1	–	+	+	–	0	0
R_2	+	–	–	+	0	0
R_3	+	+	0	0	–	+
Binary combinations of input states			Anti-symmetric sums are classical states		Symmetric sums are superposition states	
			$A_1 = R_1 - R_2$	$A_0 = R_2 - R_1$	$A_+ = R_0 - R_3$	$A_- = R_3 - R_0$

The four main *rows* of Table 2 show all the non-zero binary combinations of the orientations for vectors {**a0**, **a1**}. The four right *columns* show the possible expressions (of symmetric and anti-symmetric sums) with the non-zero or valid states highlighted. The vector and state names were chosen to represent the particular *spin* properties of the qubit, which acts like a redundantly coded classical bit with complementary states $A_0 = -A_1$. State A_0 is selected when coefficient c_0 for vector **a0** is $c_0 = +$ and state A_1 when the coefficient c_1 for **a1** is $c_1 = +$, where $c_0 = -c_1$ Because of these properties A_0 and A_1 are called the *classical states* of the qubit. Similarly, state A_+ is defined when vector coefficients $c_0 = c_1 = +$ and state A_- when vector coefficients $c_0 = c_1 = -$, thereby representing the *superposition* states (where $A_- = -A_+$). Figure 2 graphically illustrates these redundantly coded vector and state relationships.

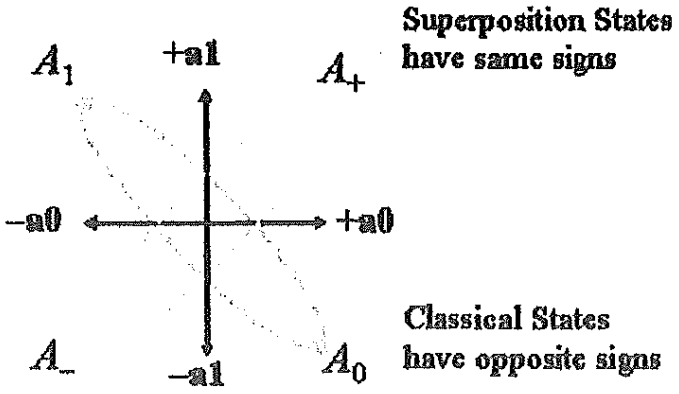


Figure 2. Vectors and States for qubit A = ±a0 ±a1

It is evident from Figure 2 that the two pairs of states $\{A_0, A_1\}$ and $\{A_-, A_+\}$ are compound states that can be represented as a vector (or line) thru the origin, but at a 45 degree angle to either axis. Therefore the sum of vectors also acts like a redundantly encoded vector, because it represents two complementary states. Because of the redundancy, there is more than one way to represent the one classical bit's worth of states in a qubit. Since the two pairs of states in Figure 2 are 90 degrees apart (which means orthogonal), they are called *out of phase* representation choices. From the physics perspective, state **a0** is the spin up state (and –**a0** means NOT **a0**) while state

a1 is the spin down state (and **–a1** means NOT **a1**). The two classical states $\{A_0, A_1\}$ represent a symmetrical spinning top pointing up or down. The two superposition states $\{A_-, A_+\}$ act like a horizontal gyroscopic top supported on one end, so is simultaneously in both/neither of the up/down states. In quantum computing, each *spin state is represented as a vector*, whereas in classical computing each *bit is represented as a vector.*

The next topic is the operator that switches between classical and superposition *states or phases*, which requires multiplication. Table 3 defines the conventional scalar multiplication table for the preceding ternary values. The terms *multiplication* and *product* are overloaded in physics and mathematics, because products exist not only for scalars, but also for vectors: the inner products, outer products, tensor products and cross products. Not to be outdone, the primary multiplication operator of geometric algebra is called the *geometric product.* The geometric product of a state and an operator (applied on the right) produces a new state, where *both states and operators are geometric algebra expressions.*

Table 3. Scalar Multiplication table for a * b = b * a

a * b	b = 0	b = +1	b = –1
a = 0	0	0	0
a = +1	0	+1	–1
a = –1	0	–1	+1

Scalar multiplication is straightforward[i] and the highlighted cells in Table 3 represent the non-zero binary combinations of the vector orientations. Those cells are equivalent to the XNOR (exclusive NOR) logic behavior, which is summarized as: *like states produce* +1 *and differing states produce* –1. XOR/XNOR based logic is identical to the odd/even parity operators and a direct result of the multiplication operator being related to XOR is the unexpected multiplicative inverse property: 1/a = a (when $a \neq 0$). This property is true for both scalars and vectors. As will be shown next, vector multiplication is slightly more complicated in geometric algebra but this complexity enables much simplicity elsewhere.

3.1. *Geometric Product and Graded N-vectors*

The geometric product can now be defined for the multiplication of vectors. As the name implies, the *geometric product* is based on topological principles. The first simple premise is that multiplying two vectors (**a b**) produces an area-like object called a *bivector*, which is a different mathematical object type than either a scalar or a vector. Multiplying 3 vectors together (**a b c**) produces a volume-like object called a *trivector*. This is easy to understand by realizing that a scalar is a grade-0 object (denoted as $\langle A \rangle_0$), a vector is a grade-1 object $\langle A \rangle_1$, a *bivector* is a grade-2 object $\langle A \rangle_2$, a *trivector* is a grade-3 object $\langle A \rangle_3$, and in general an n-vector is a grade-n object $\langle A \rangle_n$ that defines an n-volume. Adding different grade objects creates a *multivector* of

[i] Scalar multiplication is naturally closed over the ternary values {0, –1, +1}

the form; $A = \langle A \rangle_0 + \langle A \rangle_1 + \langle A \rangle_2 + ... + \langle A \rangle_n$. A *geometric algebra* G_n *spanned* by *n* orthonormal vectors contains N = 2^n unique graded elements found by expanding the expression (1+**a**)(1+**b**)... and defines 3^N unique multivectors, (i.e. $G_2 \Rightarrow 3^4 = 81$). In our definition, a qubit is a *multivector*: the sum of both grade-1 vectors in G_2.

Any *bivector* has an orientation coefficient just as a vector expression, but with the unusual geometric product identity **a b** = – **b a** This property means that the geometric product is *not commutative* (more precisely, it is anticommutative) and is simply the algebraic expression of the right-hand rule used in physics. The bivector orientation coefficient can be imagined as the right-hand thumb pointing to the front (or back) of a plane defined by a piece of paper and is depicted in Figure 3. The orientation is defined as the coefficient of any n-vector product in any grade space[i], so is equivalent to the parity of the vectors of the nvector in a particular order. This article places the vectors in standard alphabetically sorted order. Since the geometric product[ii] does not have an explicit operator, writing the product (**a b**) therefore means (**a** GP **b**), where the parentheses are optional.

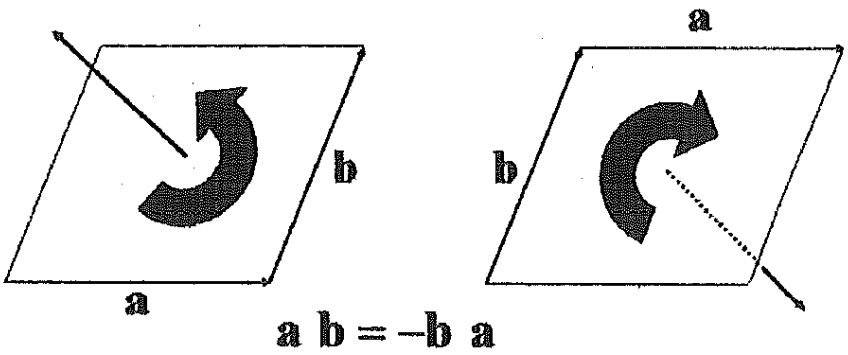


Figure 3. Geometric product is anticommutative

3.2. *Geometric Product and Spinor Operator*

The following examples demonstrate the anticommutative geometric product. Assume that a system is defined or *spanned* by a set of orthonormal vectors: G_2 = span{**a**, **b**}. Now multiply vector **a** times bivector (**a b**) and use the topological simplifications[iii] **a a** = **b b** = +1.

$$\mathbf{a}\,(\mathbf{a}\,\mathbf{b}) = \mathbf{a}\,\mathbf{a}\,\mathbf{b} = \mathbf{b} \tag{2}$$

Similarly, multiply vector **b** times bivector (**a b**) and then repeatedly multiply result by (**a b**):

$$\begin{aligned} \mathbf{b}\,(\mathbf{a}\,\mathbf{b}) &= \mathbf{b}\,\mathbf{a}\,\mathbf{b} = -\mathbf{a}\,\mathbf{b}\,\mathbf{b} = -\mathbf{a} \\ -\mathbf{a}\,(\mathbf{a}\,\mathbf{b}) &= -\mathbf{a}\,\mathbf{a}\,\mathbf{b} = -\mathbf{b} \\ -\mathbf{b}\,(\mathbf{a}\,\mathbf{b}) &= -\mathbf{b}\,\mathbf{a}\,\mathbf{b} = +\mathbf{a}\,\mathbf{b}\,\mathbf{b} = \mathbf{a} \end{aligned} \tag{3}$$

---

[i] Due to the outer product, and equivalent to the vector cross product only in three dimensions

[ii] Geometric product of vectors is the sum $\mathbf{a}\,\mathbf{b} = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\wedge\mathbf{b}$ of inner $\mathbf{a}\cdot\mathbf{b}$ and outer $\mathbf{a}\wedge\mathbf{b}$ products.

[iii] Simplification of (**a a**)=1 means a vector is self collinear and this represents the inner product $\mathbf{a}\cdot\mathbf{a} = \mathbf{1}$

As graphically depicted in Figure 4, the repeated geometric product application of the bivector $(\mathbf{a}\,\mathbf{b})$ *spins* any state counter-clockwise and explains why the bivector $(\mathbf{a}\,\mathbf{b})$ is referred to as a *spinor*. Multiplying by the bivector $(-\mathbf{a}\,\mathbf{b})$ spins the states in the clockwise direction, following the right-hand rule. The various qubit encodings are rotations of each other, so these states are rotationally invariant. You can now relax, since the spinor idea is the most difficult piece of physics and related mathematics in this article.

The result of a vector multiplied twice by a bivector inverts that vector, which can be analyzed from only the operator perspective by squaring the operator and simplifying.

$$(\mathbf{a}\,\mathbf{b})^2 = (\mathbf{a}\,\mathbf{b})(\mathbf{a}\,\mathbf{b}) = \mathbf{a}\,\mathbf{b}\,\mathbf{a}\,\mathbf{b} = -\mathbf{a}\,\mathbf{a}\,\mathbf{b}\,\mathbf{b} = -1 \tag{4}$$

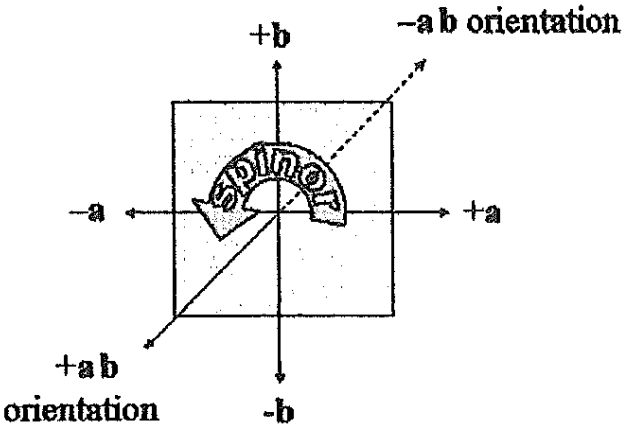


Figure 4. Bivector $(\mathbf{a}\,\mathbf{b})$ spins the state space counter-clockwise

Since $(\mathbf{a}\,\mathbf{b})^2 = -1$, therefore the spinor $\mathbf{S} = (\mathbf{a}\,\mathbf{b}) = \sqrt{-1}$. Because the spinor operator squared is the inverter (or $\mathbf{S}^2$ = NOT operator) the spinor operator is referred to as the *square root of not*: $\mathbf{S} = (\mathbf{a}\,\mathbf{b}) = \sqrt{NOT}$. This is topologically easy to understand using Figure 4 and the anticommutative geometric product[i]. Notice that even though all examples up to now are integer coefficients for spinors, it is possible to encode arbitrary angles using the GA rotators. This is important to allow arbitrary phase qubits and probability amplitudes, but will not be discussed virtually at all in this introductory paper. Another alternative is to use Tom Etter's link theory to count the multiple discrete ways of reaching the same output state.

3.3. *Reversibility, Unitary Operators, Phases and Pauli Operators*

The bivector *spinor* operator $\mathbf{S}_A = (\mathbf{a}0\,\mathbf{a}1)$ for a qubit $A = (\pm\mathbf{a}0 \pm \mathbf{a}1)$ is simply an even grade operator that switches between the odd grade classical and superposition phases[ii] and can be applied to any state. Also the inversion operator $(\mathbf{S}_A)^2 = -1$ can be applied to any state. Both the

[i] Due to these spinor properties, qubits are referred to as possessing spin ½.
[ii] A spinor is the same as the Hadamard operator.

inverter $-1 \in \langle A \rangle_0$ and the spinor $S_A \in \langle A \rangle_2$ are even grade operators, but they also have another important property, called *reversibility*. Classically speaking, a *one-to-one* mapping of states is *often* reversible but any *many-to-one* state mapping is *irreversible*.

Just as the name suggests, reversibility refers to an operator that can be *reversed or undone*. Conventional classical computing, with traditional Boolean logic gates, is typically not reversible due to many-to-one state mappings (effectively, the arrival path is lost), which means information is erased and energy is consumed due to this erasure [7]. Classical computation is reversible by using only the 3-input and 3-output reversible Toffoli or Fredkin gates, rather than the conventional irreversible 2-input gates of NAND/NOR.

Reversibility [8] is easy to describe mathematically with the understanding that all operators are implemented as products. Let's assume a multivector system state X and a multivector operator Y forming some new multivector system state $Z = X\,Y$. To *undo* this operator means convert the state Z back into state X. This is possible by simply *dividing by* Y (or multiplying times $1/Y = Y^{-1}$) resulting in $Z/Y = X\,Y/Y = X$. The operator Y is *reversible if and only if* the *multiplicative inverse* $W = 1/Y = Y^{-1}$ *exists*. An operator Y with this property (i.e. $1/Y$ exists) is called *unitary* because $Y\,W = Y\,Y^{-1} = +1$ and this formal definition is semantically synonymous with any reversible operator Y.

The good news about reversibility is that scalars ($1/a = a$), vectors ($1/\mathbf{a} = \mathbf{a}$), n-vectors ($1/S_A = -S_A$) and many multivectors ($1/A_0 = A_1$, $1/A_- = A_+$) are reversible because *geometric products are invertible*[i] [3]. The term *invertible* means (to physicists) that expressions have a *multiplicative inverse*. This term should not be confused with the similar sounding *logical inverse*, which is implemented in geometric algebra as the *additive inverse* (or negation).

A useful multivector example $P_A = -1 + S_A$, is invertible ($1/P_A = 1 + S_A$) and has several other properties. First, P_A is of even grade just like its additive operands. Second, $(P_A)^2 = S_A$ so consequently $\sqrt{S_A} = \pm P_A$. It is now possible to summarize the previously seen discrete phase relationships: $+1 = 360°$, NOT $= \sqrt{+1} = 180°$, spinor $= \sqrt{NOT} = 90°$, so the *square root* of an operator (if it exists) is related to dividing the spin angle in half and similarly the third root angle is $360°/3 = 120°$, etc. Third, since the operator -1 means inversion and S_A means phase spin, then with our interpretation of addition, P_A is *simultaneously an inversion and phase shift*! Here are the results of the *Pauli operator* P_A applied to the four qubit states.

$$\begin{aligned} A_0\,P_A = A_0(-1) + A_0\,S_A = A_1 + A_+ = (-\mathbf{a0} + \mathbf{a1}) + (\mathbf{a0} + \mathbf{a1}) = -\mathbf{a0} + \mathbf{a0} + \mathbf{a1} + \mathbf{a1} = -\mathbf{a1} \\ A_1\,P_A = A_0 + A_- = +\mathbf{a1} \\ A_-\,P_A = A_+ + A_0 = -\mathbf{a0} \\ A_+\,P_A = A_- + A_1 = +\mathbf{a0} \end{aligned} \tag{5}$$

The Pauli operator P_A *reversibly* maps the classical states $A_{1/0}$ to the vertical vector $\pm\mathbf{a1}$ and maps the superposition states $A_\pm$ to the horizontal vector $\pm\mathbf{a0}$. As expected and as graphically

[i] Neither the inner nor outer products are invertible by themselves

seen in Figure 5, this represents a discrete 45 degree phase encoding away from the classical and superposition axes. The inversion and spinor operators still function as expected for this representation. This vector encoding[i] emphasizes the classical/superposition meaning of the vectors rather than the spin up/down meaning, yet both interpretations are valid.

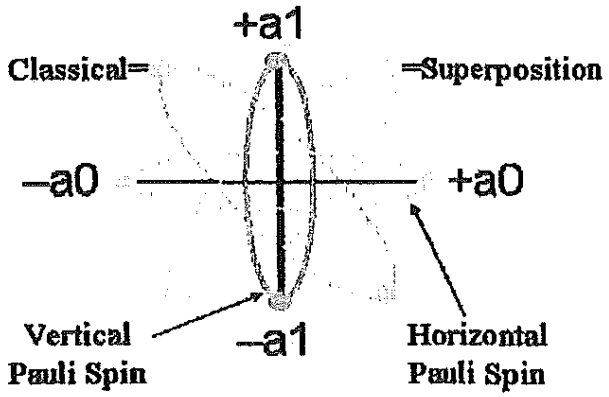


Figure 5. Phase Encodings of 180°, 90° and 45° for qubit $A = (\pm\mathbf{a0}\pm\mathbf{a1})$

All three Pauli operators have now been discussed ($\sigma_1 = -1$, $\sigma_3 = \mathbf{S}_A$, $\sigma_2 = \boldsymbol{P}_A$). This is important to quantum computing mathematics because the Pauli operators represent the reversible even-grade operators that encode how noise can affect a qubit state as either a bit inversion, a phase shift, or both a bit & phase shift *simultaneously*. Likewise, the odd-grade reversible operators $\pm\mathbf{a0}$, $\pm\mathbf{a1}$ and $(\pm\mathbf{a0}\pm\mathbf{a1})$ also produce alternate encodings to the even grade plane[ii] formed by the axes ± 1 and $\pm\mathbf{S}_A$. Quantum computation primarily involves reversibly *rotating* a qubit encoding through a phase angle[iii] without erasing the bit of information stored in the qubit, so all single-qubit operators are specific kinds of phase gates.

Of the total of $3^N - 1 = 80$ possible qubit multivectors (excludes state 0), 48 are reversible because they are invertible[iv]. The remaining 32 multivectors do not have multiplicative inverses and are thus *irreversible*. The next section describes how to identify these *irreversible* operators. Of the 80 possible multivector states, 40 multivectors are the additive inverses of the other 40, and all of these 40 unique states are discussed as sets in this article.

3.4. *Irreversibility, Singular Operators, Erasure and Measurement*

Irreversible operators are important in quantum mechanics because they erase the information encoded in a qubit. Losing information is bad, because the wrong answer will emerge when asking for an answer with a measurement question. This situation is problematic for qubits because noise is equivalent to an unwanted operator. If the system is in an "unexpected" state, then the basis-based question asked (see below) will ipso facto be ill-formed, resulting in a random binary answer from the measurement. Additionally, all measurement operators are irreversible and destroy the qubit state by setting the qubit to the questioned state. Extracting the

---

[i] "Encoding" means *basis*: classical=*standard* basis, superposition=*dual* basis, and *circular* basis= $\pm 1 \pm \mathbf{S}_A$

[ii] Identical to real (scalars) and imaginary axes ($\mathbf{S}_A = i$) as represented in complex numbers.

[iii] Geometric algebra rotators $\mathbf{a}' = R\mathbf{a}\tilde{R}$: with $R = \alpha - \beta\mathbf{e}_1\mathbf{e}_2$, $\tilde{R} = \alpha + \beta\mathbf{e}_1\mathbf{e}_2$, $\alpha = \cos(\theta/2), \beta = \sin(\theta/2)$

[iv] Definition of unitary is $|\det(X)| = +1$, which is true for all non-singular multivectors X if $\det(X) \neq 0$ [9]

information stored in a qubit destroys the state, so there is only one chance. This is similar to old core memory systems, which required writing the data back after a destructive read.

The simplest non-invertible multivector has the form $X = (\pm 1 \pm \mathbf{x})$, where $\mathbf{x}$ is any 1-vector. Simply stated, because $X^{-1} = (\pm 1 \pm \mathbf{x})^{-1}$ does not exist, then X is called *singular* [i]. This fact is the basis for all other singular operators of a qubit because using the product $X\,Y = Z$, if either factor X or Y *is singular then so is* Z[ii]. Alternatively, when X is *unitary* then X^{-1} *exists* and the multivector X is *non-singular.*

All of the 32 singular multivectors of a qubit contain one of the factors $(\pm 1 \pm \mathbf{x})$, and they are: $(\pm 1 \pm \mathbf{a0}) = 4$, $(\pm 1 \pm \mathbf{a1}) = 4$, $\mathbf{a0}(\pm 1 \pm \mathbf{a1}) = 4$, $\mathbf{a1}(\pm 1 \pm \mathbf{a0}) = 4$, $(\pm 1 \pm \mathbf{a0})(\pm 1 \pm \mathbf{a1}) = 8$ and the opposite order $(\pm 1 \pm \mathbf{a1})(\pm 1 \pm \mathbf{a0}) = 8$ for a total of $4 + 4 + 4 + 4 + 8 + 8 = 32$ unique singular multivectors. As is shown below, all of these singular operators are related to measurement and information erasure. Each operator X in the list above was proved to be singular by exhaustively attempting to solve the equality $X\,Y = 1$, for each of the possible 80 multivectors Y, and no solutions were found. For expressions involving multiple qubits, other singular expressions exist, however, that do not have $(\pm 1 \pm \mathbf{x})$ as a factor.

Knowing exactly how measurement occurs, answers are extracted, and information is erased, in a qubit is important for quantum computing, and singular operators are an important clue to this understanding. Essentially, a measurement entails asking what state orientation a particular vector currently possesses. In geometric algebra, a multivector of the form $X = (-1)(1 \pm \mathbf{x})$ can be used to isolate only the state cases for orientation $\pm\mathbf{x}$ and so is equivalent to *testing* or *decoding* vector $\mathbf{x}$ for a particular orientation, denoted as $X_{\pm}$. Each of the four output columns in Table 4 represents one of the singular expressions of the form $(-1)(1 \pm \mathbf{x})$.

In every column, two rows contain the + state and two rows contain the 0 state. When this expression is used as an operator it effectively creates a notch filter that only passes the non-zero states. By combining two orientation choices using the geometric product, a particular row can be selected, which specifies the logically combined state $A0_{\pm}$ *and* $A1_{\pm}$, so each row R_k represents a cell in a Boolean logic Karnaugh map[iii] used by conventional logic designers.

Table 4. Specifying a particular vector orientation in $G_2 = \text{span}\{\mathbf{a0}, \mathbf{a1}\}$.

Row_k	a0	a1	$(-1)(1 - \mathbf{a0})$	$(-1)(1 + \mathbf{a0})$	$(-1)(1 - \mathbf{a1})$	$(-1)(1 + \mathbf{a1})$
R_0	–	–	+	0	+	0
R_1	–	+	+	0	0	+
R_2	+	–	0	+	+	0
R_3	+	+	0	+	0	+
Summation of R_k ➔			$A0_- = R_0 + R_1$	$A0_+ = R_2 + R_3$	$A1_- = R_0 + R_2$	$A1_+ = R_1 + R_3$
Denoted as Vector[iv] ➔			[+ + 0 0]	[0 0 + +]	[+ 0 + 0]	[0 + 0 +]

[i] X is singular if $\det(X) = 0$ because X^{-1} becomes *infinite* due to X^{-1} being dependent on $1/\det(X)$

[ii] For $X\,Y = Z$, then $\det(X)\det(Y) = \det(Z)$, so if $\det(X) = 0$ or $\det(Y) = 0$ then $\det(Z) = 0$

[iii] R_k = *computational* basis: different than *standard* since mult=XNOR vs. AND in Hilbert spaces

[iv] The vector notation is set of R_k denoted as vector $[R_0\ R_1\ R_2\ R_3\ \ldots]$ and is used extensively here.

$A0_{\pm}\, A1_{\pm} = (-1)(1 \pm \mathbf{a0})(-1)(1 \pm \mathbf{a1}) = (1 \pm \mathbf{a0})(1 \pm \mathbf{a1}) = (1 \pm \mathbf{a0} \pm \mathbf{a1} \pm \mathbf{a0\ a1})$, whence

$$A0_{-}\, A1_{-} = (1 - \mathbf{a0} - \mathbf{a1} + \mathbf{a0\ a1})$$
$$A0_{-}\, A1_{+} = (1 + \mathbf{a0} - \mathbf{a1} - \mathbf{a0\ a1})$$
$$A0_{+}\, A1_{-} = (1 - \mathbf{a0} + \mathbf{a1} - \mathbf{a0\ a1})$$
$$A0_{+}\, A1_{+} = (1 + \mathbf{a0} + \mathbf{a1} + \mathbf{a0\ a1}) \qquad (6)$$

Table 5. Specifying two vector orientations in G_2

Row k	a0	a1	(1–a0)(1–a1)	(1–a0)(1+a1)	(1+a0)(1–a1)	(1+a0)(1+a1)
R_0	–	–	+	0	0	0
R_1	–	+	0	+	0	0
R_2	+	–	0	0	+	0
R_3	+	+	0	0	0	+
State logic ➔			$R_0 = A0_{-}\, A1_{-}$	$R_1 = A0_{-}\, A1_{+}$	$R_2 = A0_{+}\, A1_{-}$	$R_3 = A0_{+}\, A1_{+}$
Denoted as Vector ➔			$R_0 = [+\ 0\ 0\ 0]$	$R_1 = [0\ +\ 0\ 0]$	$R_2 = [0\ 0\ +\ 0]$	$R_3 = [0\ 0\ 0\ +]$

Table 5 illustrates these singular expressions, which represent the *topologically smallest features* in a qubit representation. These row-decode operators, R_k are *linearly independent* and all other expressions can be derived by *summing specific rows*, so each algebraic expression has a unique, dual, sparse representation expressed as the sum of R_k. The inverse of R_k is denoted as $P_k = -R_k$. The compact vector-like notation $[R_0\ R_1\ R_2\ R_3]$ expresses these states, where the row values $R_k \in \{0, -, +\}$ are the values of the expressions for every non-zero combination of vector orientations. This vector notation can be thought of as a matrix diagonal because $R_0+R_1+R_2+R_3 = [+\,+\,+\,+] = +1$, and $P_0+P_1+P_2+P_3 = [-\,-\,-\,-] = -1$. The vector notations for several other familiar multivectors are: $\mathbf{a0} = [-\,-\,+\,+]$, $\mathbf{a1} = [-\,+\,-\,+]$, $\mathbf{S}_A = [+\,-\,-\,+]$, $A_0 = [0\,-\,+\,0]$, $A_1 = [0\,+\,-\,0]$, $A_+ = [+\,0\,0\,-]$, $A_- = [-\,0\,0\,+]$ and $\boldsymbol{P}_A = [0\,+\,+\,0]$. Element by element vector addition is identical to algebraic addition, for example the sum: $\mathbf{a0} + \mathbf{a1} = [-\,-\,+\,+] + [-\,+\,-\,+] = [+\,0\,0\,-] = A_+$, because the R_k are linearly independent.

The overall qubit singular-operator relationships are now shown in Table 6, which illustrates the *answer* to measuring the four qubit states (in first column) from the perspective of each singular row-decode operator $R_k = A0_{\pm}\, A1_{\pm}$. This table is an example of a set of one-to-one mappings that is *irreversible* because the mapping operators are singular and so cannot be undone. Classical Boolean logic systems do not have the concept of singular operators.

Table 6. Qubit measurement results for G_2

Start States A	Each start state A times each R_k			
	A(1+a0)(1–a1)	A(1–a0)(1+a1)	A(1+a0)(1+a1)	A(1–a0)(1–a1)
$A_0 = +\ \mathbf{a0} - \mathbf{a1}$	$-1 + \mathbf{a1} = \boldsymbol{I}^+$	$+1 + \mathbf{a1} = \boldsymbol{I}^-$	$-\mathbf{a0}\,(+1 + \mathbf{a1})$	$+\mathbf{a0}\,(-1 + \mathbf{a1})$
$A_1 = -\ \mathbf{a0} + \mathbf{a1}$	$+1 - \mathbf{a1} = \boldsymbol{I}^-$	$-1 - \mathbf{a1} = \boldsymbol{I}^+$	$-\mathbf{a0}\,(-1 - \mathbf{a1})$	$+\mathbf{a0}\,(+1 - \mathbf{a1})$
$A_- = -\ \mathbf{a0} - \mathbf{a1}$	$-\mathbf{a0}\,(-1 + \mathbf{a1})$	$+\mathbf{a0}\,(+1 + \mathbf{a1})$	$+1 + \mathbf{a1} = \boldsymbol{I}^-$	$-1 + \mathbf{a1} = \boldsymbol{I}^+$
$A_+ = +\ \mathbf{a0} + \mathbf{a1}$	$-\mathbf{a0}\,(+1 - \mathbf{a1})$	$+\mathbf{a0}\,(-1 - \mathbf{a1})$	$-1 - \mathbf{a1} = \boldsymbol{I}^+$	$+1 - \mathbf{a1} = \boldsymbol{I}^-$
End State ➔	$A \Rightarrow +\ \mathbf{a0} - \mathbf{a1}$	$A \Rightarrow -\ \mathbf{a0} + \mathbf{a1}$	$A \Rightarrow +\ \mathbf{a0} + \mathbf{a1}$	$A \Rightarrow -\ \mathbf{a0} - \mathbf{a1}$
Description ➔	Classical States Measurement		Superposition States Measurement	

Applying the singular operators R_k, Table 6 produces two kinds of singular *answers*, either a "sparse invariant" or a random value. The measurement returns the answer and the qubit changes to the *end state* after measurement. The resulting answers of the form $(\pm 1 \pm \mathbf{a1}) = I^{\pm}$ *act like a constant* since the *non-zero output row-states* are either all + or all –, as follows. This was originally hinted at in Table 4, since $-1 + \mathbf{a1} = [+\,0\,+\,0] = I^{+}$ and $-1 - \mathbf{a1} = [0\,+\,0\,+] = I^{+}$ are two out-of-phase examples of *sparse invariants*. This name was coined because the multivectors $I^{\pm}$ act like sparse versions of the constants ±1, with the properties $I^{-} = -I^{+}$ and $(I^{\pm})^2 = I^{+}$. The sum of two out-of-phase versions of these invariants form the constants $+1 = I^{+}_{0^\circ} + I^{+}_{90^\circ} = [+\,+\,+\,+]$ and $-1 = I^{-}_{0^\circ} + I^{-}_{90^\circ} = [-\,-\,-\,-]$. Any multivector of the form $(\pm 1 \pm X)$ is a sparse invariant, where X is any n-vector. Not all sparse invariants are singular, e.g. $P_A = -1 + S_A = [0 + + 0]$.

From a measurement perspective, the sparse invariants $I^{\pm}$ *represent a Boolean answer* because the result is I^{+} or I^{-}, and the qubit is *projected* to the end state matching the question. This process is irreversible because both R_k and $I^{\pm}$ are singular. From the sums of R_k or vector notation, it is easy to see how information is erased because the *symmetry*[i] *of the qubit is broken.* The symmetry is essentially based on which rows are valid, where the rows $\{R_1, R_2\}$ are non-zero only for the classical states and the rows $\{R_0, R_3\}$ are non-zero only for the superposition states[ii]. The sparse invariants include a row state from each pair of rows $I^{+}_{0^\circ} = [+\,0\,+\,0] = R_0 + R_2$ and $I^{+}_{90^\circ} = [0\,+\,0\,+] = R_1 + R_3$, so the combined *asymmetrical state* is *no longer linearly independent* since it is the sum of non-orthogonal elements[iii].

The row-pair symmetry is also broken by singular operators of the form $(\pm \mathbf{a0} \pm \mathbf{a0}\ \mathbf{a1})$ because $\mathbf{a0} + \mathbf{a0}\ \mathbf{a1} = R_1 - R_3$, $-\mathbf{a0} - \mathbf{a0}\ \mathbf{a1} = R_3 - R_1$, $\mathbf{a0} - \mathbf{a0}\ \mathbf{a1} = R_0 - R_2$, and $-\mathbf{a0} + \mathbf{a0}\ \mathbf{a1} = R_2 - R_0$. Each of these results looks like a random value because half the states are + and other half are –, or *statistically random*, in contrast to the invariants, which are all the same value. The row-decode operators $R_k = A0_{\pm}\, A1_{\pm}$ are also asymmetrical since they each contain only one non-zero row.

The above discussion utilizes only half of the singular states of a qubit. Exactly the same analysis can be performed using the *anticommutative* or dual versions of the row-decode operator products $R_{7\text{-}k} = A1_{\pm}\, A0_{\pm}$ (dual of $R_k = A0_{\pm}\, A1_{\pm}$). These expressions represent the other four multivectors of the form $(1 \pm \mathbf{a0} \pm \mathbf{a1} \pm \mathbf{a0}\ \mathbf{a1})$, where the sign is inverted for the bivector, resulting in all zero-valued row-states being converted to the – state.

$$
\begin{aligned}
A1_{+}\, A0_{+} &= (1 + \mathbf{a0} + \mathbf{a1} - \mathbf{a0}\ \mathbf{a1}) = [+\,-\,-\,-] = R_7 \text{ where } R_0 = [+\,0\,0\,0] \\
A1_{-}\, A0_{+} &= (1 + \mathbf{a0} - \mathbf{a1} + \mathbf{a0}\ \mathbf{a1}) = [-\,+\,-\,-] = R_6 \text{ where } R_1 = [0\,+\,0\,0] \\
A1_{+}\, A0_{-} &= (1 - \mathbf{a0} + \mathbf{a1} + \mathbf{a0}\ \mathbf{a1}) = [-\,-\,+\,-] = R_5 \text{ where } R_2 = [0\,0\,+\,0] \\
A1_{-}\, A0_{-} &= (1 - \mathbf{a0} - \mathbf{a1} - \mathbf{a0}\ \mathbf{a1}) = [-\,-\,-\,+] = R_4 \text{ where } R_3 = [0\,0\,0\,+]
\end{aligned}
\qquad (7)
$$

[i] Symmetry or coherence, whereas asymmetry means decoherence

[ii] Pair-wise orthogonal $R_1 \circ R_2 = 0$ are the standard basis and $R_0 \circ R_3 = 0$ are the dual basis.

[iii] Non-orthogonal vectors cannot be used as the matrix basis vectors for quantum systems.

With the inverted operators $P_{7-k} = -R_{7-k}$ also defined, then the following facts are true about R_{4-7}: $R_4+R_5+R_6+R_7 = +1$ and $P_4+P_5+P_6+P_7 = -1$. The overall *unitarity*[i] property of a qubit is defined as $P_0+P_1+P_2+P_3+P_4+P_5+P_6+P_7 = +1$ and $R_0+R_1+R_2+R_3+R_4+R_5+R_6+R_7 = -1$.

An important and interesting topological fact is that these set of eight multivectors have the invertiblity property $X = 1/X = X^{-1}$, and therefore are self-unitary: $X X^{-1} = XX = X^2 = 1$. The multivectors in G_2 with this property[ii] have the form of $E_k = (\pm\mathbf{a0} \pm\mathbf{a1} \pm\mathbf{a0}\ \mathbf{a1})$ and represent the eight corners of the cube in Figure 6, formed by the axes $\{\pm\mathbf{a0}, \pm\mathbf{a1}, \pm\mathbf{a0}\ \mathbf{a1}\}$. These multivectors form the corners of the dual tetrahedrons formed by the sides $P_k = -(1+E_k)$ or $E_k = R_k - 1$ shown in Figure 7. Even though the axes are drawn in a cube, they are not orthogonal.

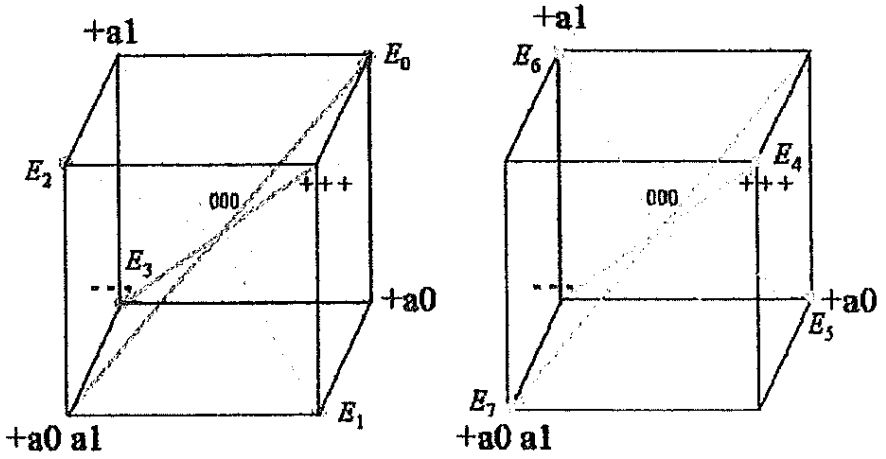


Figure 6. Eight multivectors E_k define two sets (E_{0-3} and E_{7-4}) of four corners

The results in Figures 6 and 7 are topologically interesting and very relevant to matrix mathematics. One of the important results of the relationships, $R_k = (1 + E_k)$ and $(E_k)^2 = 1$ is that the product [10] $E_k\ R_k = E_k\ (1+E_k) = E_k + (E_k)^2 = E_k + 1 = R_k$, which ultimately leads to the important result[iii] that $P_k\ P_k = (P_k)^2 = P_k$, where the P_k form the sides of the dual tetrahedrons in Figure 7. Table 7 summarizes these multivector relationships including the sum of all $E_k = 0$.

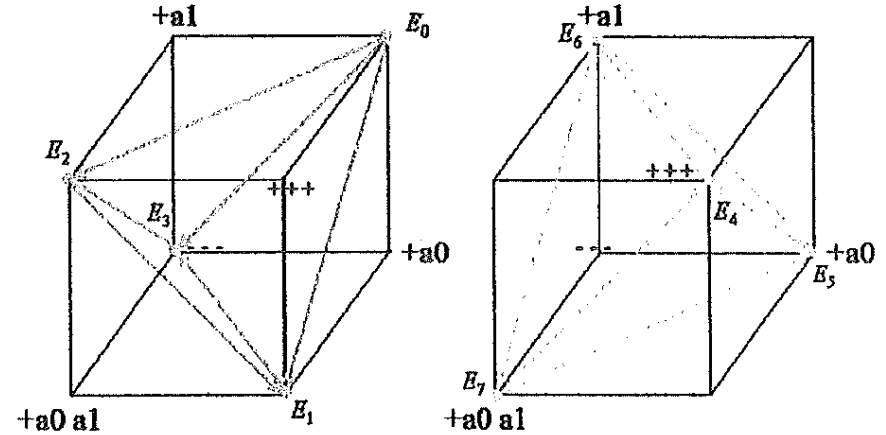


Figure 7. Sides of a tetrahedron are formed by P_{0-3} on left and P_{7-4} on right

---

[i] Same as the unitarity constraint for qubits in Hilbert Space

[ii] Property $E_k\ E_k = 1$ means the E_k are the eigenvectors and $P_k = -(1+E_k)$ are the projection operators

[iii] P_k are *idempotent* $(P_k)^2 = P_k$ projection operators of a qubit. P_k are eigenvalues of the eigenvectors E_k

The symmetric results in Table 7 show that our algebraic notation naturally describes a qubit and is formally equivalent to the matrix notation traditionally used for the same purpose. Even though *establishing* the foundational concepts of qubits relies on some fairly abstruse mathematics, once these are in place, one need only the relatively straightforward manipulation of geometric algebra to read, write, manipulate, interpret, and understand qubits. Nevertheless, quantum concepts themselves still constitute a relatively steep learning curve.

Table 7. Summary of Definitions and Relationships between R_k, P_k and E_k

Primary Tetrahedron				Dual Tetrahedron			
k =	$E_k = R_k - 1$	$P_k = -R_k$	$R_k = 1 + E_k$	k =	$E_k = R_k - 1$	$P_k = -R_k$	$R_k = 1 + E_k$
0	[0 − − −]	[− 0 0 0]	[+ 0 0 0]	7	[0 + + +]	[− + + +]	[+ − − −]
1	[− 0 − −]	[0 − 0 0]	[0 + 0 0]	6	[+ 0 + +]	[+ − + +]	[− + − −]
2	[− − 0 −]	[0 0 − 0]	[0 0 + 0]	5	[+ + 0 +]	[+ + − +]	[− − + −]
3	[− − − 0]	[0 0 0 −]	[0 0 0 +]	4	[+ + + 0]	[+ + + −]	[− − − +]
sum	[0 0 0 0]	[− − − −]	[+ + + +]	sum	[0 0 0 0]	[− − − −]	[+ + + +]

The last remaining set of expressions from the 80 qubit states[i] is called the *trine* states. Trines are mathematically easy to identify because they represent the eight solutions of the equality $(Tr)^3 = 1$. The qubit solutions all have the form $Tr = (+1 \pm \mathbf{a0} \pm S_A)$ or $Tr = (+1 \pm \mathbf{a1} \pm S_A)$ and their inverses. The general form is the concurrent sum of the spinor and a singular operator of the form $(+1 \pm \mathbf{x})$. As expected and as seen in state evolution in Eq. (8), this 120° operator causes the state space to become asymmetrical. These operators are *unitary* though, because the multivector Tr is invertible since $1/Tr = (Tr)^2$.

$$\begin{gathered} A_0 = [0 + - 0] \\ A_0\,(+1 + \mathbf{a0} + S_A) = (+1 - \mathbf{a0} + S_A) = [0 + - +] \\ A_0\,(+1 + \mathbf{a0} + S_A)^2 = (-1 + \mathbf{a0} - S_A) = [0 + - -] \\ A_0\,(+1 + \mathbf{a0} + S_A)^3 = A_0 = [0 + - 0] \end{gathered} \tag{8}$$

The next section describes combining multiple qubits to form a quantum register.

4. Quantum Registers as Geometric Product of Qubits

Multiple q qubits can be combined to form a quantum register $Q_q = G_{n=2q}$ that defines a space of size $n = 2q$. The state space of two qubits[ii] with $n = 4$ does not have the size of 4 + 4 = 8 states, but rather $N = 2^4 = 16 = 4 * 4$ total states and $3^{16} = 43{,}046{,}721$ discrete multivectors. The number of states grows exponentially because combining qubits entails *entangling* their state spaces. Geometric algebra easily expresses qubit *entanglement* using the geometric product[iii]. The entanglement of $q = 2$ qubits, defined as $A = (\pm\mathbf{a0} \pm \mathbf{a1})$ and $B = (\pm\mathbf{b0} \pm \mathbf{b1})$, is simply the geometric product $A\,B$ of the qubits:

$$A\,B = (\pm\mathbf{a0} \pm \mathbf{a1})(\pm\mathbf{b0} \pm \mathbf{b1}) = \pm\,\mathbf{a0\,b0} \pm \mathbf{a0\,b1} \pm \mathbf{a1\,b0} \pm \mathbf{a1\,b1} \tag{9}$$

---

[i] For the full table of 40/80 operators see table 7.2 in reference [4].

[ii] $G_{n=3}$ is a qutrit where multivector state $A = (\pm\mathbf{a0} \pm \mathbf{a1} \pm \mathbf{a2})$ and describes a spin-one particle: a photon.

[iii] Geometric product is same as tensor product $\otimes$ in Hilbert spaces and tensor power $X^{\otimes n}$ is simply X^n

This sum of four bivectors represents all the possible simultaneous combinations of the spin vectors. Recalling the spinor notation for each qubit (i.e. S_A, S_B, etc), these bivectors are actually cross-qubit spinors and are denoted as S_{00} = **a0 b0**, S_{01} = **a0 b1**, S_{10} = **a1 b0** and S_{11} = **a1 b1**, with all vectors in the standard sorted order. The *product of sums* format on the left is mathematically identical to the *sum of products* format on the right. If a sum of bivectors can be factored back into a product of sums format, the entangled states are called *separable*.

Specific examples with each qubit in specific states produce a vector notation with 16 rows. The number of states grows as $N = 2^{2q} = 4^q$, but the number of non-zero states only grows as $2^q = 4$. Notice that sum of products for $A_0 B_1$ is *indistinguishable* from $A_1 B_0$ so $A_0 B_1 = A_1 B_0$.

$$\begin{aligned} A_0 B_0 &= (\mathbf{a0}-\mathbf{a1})(\mathbf{b0}-\mathbf{b1}) = +\mathbf{a0\ b0} - \mathbf{a0\ b1} - \mathbf{a1\ b0} + \mathbf{a1\ b1} \\ A_0 B_1 &= (\mathbf{a0}-\mathbf{a1})(\mathbf{b1}-\mathbf{b0}) = -\mathbf{a0\ b0} + \mathbf{a0\ b1} + \mathbf{a1\ b0} - \mathbf{a1\ b1} \\ A_1 B_0 &= (\mathbf{a1}-\mathbf{a0})(\mathbf{b0}-\mathbf{b1}) = -\mathbf{a0\ b0} + \mathbf{a0\ b1} + \mathbf{a1\ b0} - \mathbf{a1\ b1} \\ A_+ B_+ &= (\mathbf{a0}+\mathbf{a1})(\mathbf{b0}+\mathbf{b1}) = +\mathbf{a0\ b0} + \mathbf{a0\ b1} + \mathbf{a1\ b0} + \mathbf{a1\ b1} \end{aligned} \tag{10}$$

Using the multiplication principle 0 x = 0, then the valid or non-zero states of both qubits must be satisfied simultaneously. As shown in Table 8, if the 16 row vectors[i] are determined for the above examples, then the valid rows are: $A_0\ B_0 = -R_5 + R_6 + R_9 - R_{10}$ and $A_+ B_+ = R_0 - R_3 - R_{12} + R_{15}$ based on the simultaneity constraint that both qubits are contributing non-zero states.

Table 8. Valid rows for products $A_0 B_0$ and $A_+ B_+$ in Q_2

Row$_k$	State Combinations				Individual bivector products				Column Vector	
	a0	**a1**	**b0**	**b1**	**a0 b0**	**a0 b1**	**a1 b0**	**a1 b1**	$A_+ B_+$	$A_0 B_0$
R_0	–	–	–	–	+	+	+	+	+	0
R_1	–	–	–	+	+	–	+	–	0	0
R_2	–	–	+	–	–	+	–	+	0	0
R_3	–	–	+	+	–	–	–	–	–	0
R_4	–	+	–	–	+	+	–	–	0	0
R_5	–	+	–	+	+	–	–	+	0	–
R_6	–	+	+	–	–	+	+	–	0	+
R_7	–	+	+	+	–	–	+	+	0	0
R_8	+	–	–	–	–	–	+	+	0	0
R_9	+	–	–	+	–	+	+	–	0	+
R_{10}	+	–	+	–	+	–	–	+	0	–
R_{11}	+	–	+	+	+	+	–	–	0	0
R_{12}	+	+	–	–	–	–	–	–	–	0
R_{13}	+	+	–	+	–	+	–	+	0	0
R_{14}	+	+	+	–	+	–	+	–	0	0
R_{15}	+	+	+	+	+	+	+	+	+	0

[i] For Q_q the $P_k = -R_k$ are singular, but are idempotent only if the definition is extended to: $(P_k)^{n=2q} = P_k$

As expected, the valid states of the system are just the valid states for each qubit spread out across a larger space. The green highlighted rows $\{R_5, R_6, R_9, R_{10}\}$ indicate the classical states A_0 and B_0. The blue highlighted rows $\{R_0, R_3, R_{12}, R_{15}\}$ indicate the superposed states A_+ and B_+. A very interesting intermediate result noted in the rose colored middle columns is an output state can only be zero if the sum of 2^q bivector orientations exactly equals 0. This only occurs when all bivectors have exactly an *equal number of both orientations*[i]. Consequently, all non-zero outputs can occur only when *all the bivector orientation coefficients have exactly the same sign.* This pair-wise cancellation result is therefore independent of the mod 3 addition conventions established initially. For more examples, discussion and proof see [4].

Separable qubits each can be individually manipulated using the appropriate operators, and the operators can be thought of as being *sequentially applied*, producing various intermediate states. Due to non-commutative products, remember that $A_0\ B_0 = -B_0\ A_0$ (except for even grade operators that are commutative, such as $B\ \mathbf{S}_A = \mathbf{S}_A\ B$).

$$\begin{gathered} A_0\ B_0\ \mathbf{S}_A = A_0\ \mathbf{S}_A\ B_0 = A_+\ B_0 = +\ \mathbf{a0\ b0} - \mathbf{a0\ b1} + \mathbf{a1\ b0} - \mathbf{a1\ b1} \\ A_0\ B_0\ \mathbf{S}_B = A_0\ B_+ = +\ \mathbf{a0\ b0} + \mathbf{a0\ b1} - \mathbf{a1\ b0} - \mathbf{a1\ b1} \\ A_0\ B_0\ \mathbf{S}_A\ \mathbf{S}_B = A_0\ \mathbf{S}_A\ B_0\ \mathbf{S}_B = A_+\ B_+ = +\ \mathbf{a0\ b0} + \mathbf{a0\ b1} + \mathbf{a1\ b0} + \mathbf{a1\ b1} \end{gathered} \tag{11}$$

Also understand that the Pauli operators applied to both qubits define the cross-qubit spinors.

$$\begin{gathered} A_0\ B_0\ \mathbf{P}_A\ \mathbf{P}_B = A_0\ \mathbf{P}_A\ B_0\ \mathbf{P}_B = \mathbf{a1\ b1} = \mathbf{S}_{11} \text{ and likewise} \\ A_+\ B_+\ \mathbf{P}_A\ \mathbf{P}_B = \mathbf{a0\ b0} = \mathbf{S}_{00} \\ A_+\ B_1\ \mathbf{P}_A\ \mathbf{P}_B = \mathbf{a0\ b1} = \mathbf{S}_{01} \\ A_1\ B_+\ \mathbf{P}_A\ \mathbf{P}_B = \mathbf{a1\ b0} = \mathbf{S}_{10} \end{gathered} \tag{12}$$

This implies that the sum of spinor products is identical to representing the qubits in four distinct states simultaneously (i.e. superposed) in the Pauli encoding. In fact, this is exactly the previous meaning of a sum of cross-qubit spinors, since addition means *concurrent.*

4.1. *Ebits and Bell States*

A very interesting result regarding two qubits is applying both spinors *concurrently* $(\mathbf{S}_A + \mathbf{S}_B)$ rather than sequentially $(\mathbf{S}_A\ \mathbf{S}_B)$ to produce an *ebit.* Half of the bivectors disappear due to destructive interference. As a consequence, this result is *inseparable* and the reason is the erasure of phase-states. Just as a single qubit is a computational resource due to superposition of states, an ebit is also a computational resource because it encodes exactly one classical bit of information (one bit being erased), even if the qubits are separated by a large distance [11]. The ebit's property is that of an Einstein-Podolsky-Rosen (EPR) communications resource.

$$A_0\ B_0\ (\mathbf{S}_A + \mathbf{S}_B) = A_+\ B_0 + A_0\ B_+ = -\mathbf{a0\ b0} + 0\ \mathbf{a0\ b1} + 0\ \mathbf{a1\ b0} + \mathbf{a1\ b1} = -\mathbf{a0\ b0} + \mathbf{a1\ b1} \tag{13}$$

[i] The number of spinors $s=2^q$ contains *only even factors*, so $s/3 = \pm 1 \neq 0$, so 0 occurs only when $+1 - 1 = 0$

This state is one of the four *Bell states* [i] B_1. The *concurrent spinor* $B = (S_A + S_B)$, which turns out to be the *Bell operator*, iteratively generates all four Bell states ($B_0 \Rightarrow B_1 \Rightarrow B_2 \Rightarrow B_3 \Rightarrow B_0$) using the formula $B_{i+1} = B_i B$. Table 8 shows the very interesting result that the only valid states are where *exactly one qubit* occupies the superposition state at a time. The unlisted rows are zero, so do not occur. This property is also holds true for valid row states for any number of qubits as: $A_0 B_0 C_0 \ldots (S_A + S_B + S_C + \ldots)$. This symmetry is quite fascinating!

The even numbered Bell states are complements of each other $B_0 = -B_2$ and the same is true for the odd numbered states $B_1 = -B_3$. This suggests something about the square of the Bell operator and as expected, a higher dimensional version of the sparse invariants surfaces.

$$B\,B = (B)^2 = +1 - S_A S_B = [0- -0 -00- -00- 0- -0] = I^-$$
$$(B)^4 = -1 + S_A S_B = [0++0 +00+ +00+ 0++0] = I^+ \qquad (14)$$

Table 8. Valid rows for ebit B_0 in Q_2

Row$_k$	State Combinations				Individual bivectors		Output column
	a0	a1	b0	b1	–a0 b0	a1 b1	
R_1	–	–	–	+	–	–	+
R_2	–	–	+	–	+	+	–
R_4	–	+	–	–	–	–	+
R_7	–	+	+	+	+	+	–
R_8	+	–	–	–	+	+	–
R_{11}	+	–	+	+	–	–	+
R_{13}	+	+	–	+	+	+	–
R_{14}	+	+	+	–	–	–	+

An important question is, "Is the Bell operator singular?" The answer is yes, because $(B)^{-1}$ does not exist [4], which means that once the Bell operator is applied, the combined states cannot be exited or escaped using a unitary operator. Applying the inverted operator $-B$ evolves the states in the opposite direction $B_{i-1} = B_i(-B)$.

How the Bell operator erases information can easily be demonstrated once the magic operator and magic states are defined. The four *magic states* [ii] ($M_0 \Rightarrow M_1 \Rightarrow M_2 \Rightarrow M_3 \Rightarrow M_0$) are generated by the singular *magic operator* $M = (S_A - S_B)$ using the iteration $M_{i+1} = M_i M$. The magic states produce 90° out-of-phase sparse invariants compared to the Bell versions.

$$M\,M = (M)^2 = +1 + S_A S_B = [-00- 0- -0 0- -0 -00-] = I^-$$
$$(M)^4 = -1 - S_A S_B = [+00+ 0++0 0++0 +00+] = I^+ \qquad (15)$$

---

[i] $B_0 = -S_{00} + S_{11} = \Phi^+$, $B_1 = S_{01} + S_{10} = \Psi^+$, $B_2 = S_{00} - S_{11} = \Phi^-$, $B_3 = -S_{01} - S_{10} = \Psi^-$

[ii] $M_0 = S_{01} - S_{10}$, $M_1 = -S_{00} - S_{11}$, $M_2 = -S_{01} + S_{10}$, $M_3 = S_{00} + S_{11}$

It is possible to switch reversibly between the Bell and the magic states because $M_3 = B_2 (S_{01}+S_{10})$. An important relation for Bell and magic states is: $B_i M = M_i B = 0$, which follows from the complete destructive interference of these state and operator spinors. Armed with this knowledge, one can usefully express the original entanglement equations as the sum of Bell and magic states.

$$A\,B = (\pm\mathbf{a0}\pm\mathbf{a1})(\pm\mathbf{b0}\pm\mathbf{b1}) = \pm\,\mathbf{a0\,b0} \pm \mathbf{a0\,b1} \pm \mathbf{a1\,b0} \pm \mathbf{a1\,b1} = B_j + M_i \tag{16}$$

Some particular examples are:

$$\begin{aligned} A_0\,B_0 &= (\mathbf{a0}-\mathbf{a1})(\mathbf{b0}-\mathbf{b1}) = +\,\mathbf{a0\,b0} - \mathbf{a0\,b1} - \mathbf{a1\,b0} + \mathbf{a1\,b1} = B_3 + M_3 \\ A_+\,B_+ &= (\mathbf{a0}+\mathbf{a1})(\mathbf{b0}+\mathbf{b1}) = +\,\mathbf{a0\,b0} + \mathbf{a0\,b1} + \mathbf{a1\,b0} + \mathbf{a1\,b1} = B_1 + M_3 \end{aligned} \tag{17}$$

Therefore, independent of the starting state, half of the states are *always multiplicatively erased* when applying either the Bell or magic operators because $B_i M = M_i B = 0$. These results show that information is erased and these operators are irreversible, since a many-to-one mapping occurs due to erasure, as illustrated with the examples $A_0 B_0 M$ and $A_+ B_+ M$:

$$\begin{aligned} A_0\,B_0\,B &= B_0 + 0 \text{ and } A_0\,B_0\,M = 0 + M_0 \\ A_+\,B_+\,B &= B_2 + 0 \text{ and } A_+\,B_+\,M = 0 + M_0 \end{aligned} \tag{18}$$

A simple proof that B and M are singular can also be realized using the Cancellation Principle of Multiplication of multivectors which states: if $X\,Y = X\,Z$ then $Y = Z$ *if and only if* $1/X$ exists. The proof uses an example: if $X = Y = B$ and $Z = P_A\,P_B\,(-1)$, it can be shown that:

$$\begin{aligned} B\,B &= B\,P_A\,P_B\,(-1) = 1 - S_A S_B \text{ is } \textit{always True} \text{ but} \\ Z &= P_A\,P_B\,(-1) = -1 + S_A + S_B - S_A S_B = B - (1 + S_A S_B) \end{aligned} \tag{19}$$

The equality $B = B - (1 + S_A S_B)$ can be true only if $(1 + S_A S_B) = 0$, which is always False even though the product $B(1 + S_A S_B) = 0$ is always True. This contradiction therefore means $B \neq B - (1 + S_A S_B)$ because $1/B$ does not exist[i] and B is singular. Similarly, M is singular.

The Bell and magic states can also be used as singular operators[ii] to *orient* the states, because: $B_i\,B_i = I^-$, $B_i\,B_{i+2} = I^+$ while $B_i\,B_{i+1} = B_i\,B_{i-1} =$ *random states*, and likewise for M_i. See Figure 8 for a graphical summary of the states, where $P_{AB} = P_A\,P_B$. It is easy to understand that for three (or more) qubits, there are $(q-1)^2 = 4$ equivalent Bell operators of the form $(S_A \pm S_B \pm S_C)$ and the same number of out-of-phase sets of Bell states with exactly the same properties discussed here. This concludes the discussion of ebits and the Bell and magic states. The next topic is the new operators that are possible for two qubits.

---

[i] Exhaustively searched the 43 million cases for solutions X in Q_2 where $(S_A \pm S_B)(X) = 1$ and found none.
[ii] All B_i and M_i are singular because they respectively contain B and M as factors.

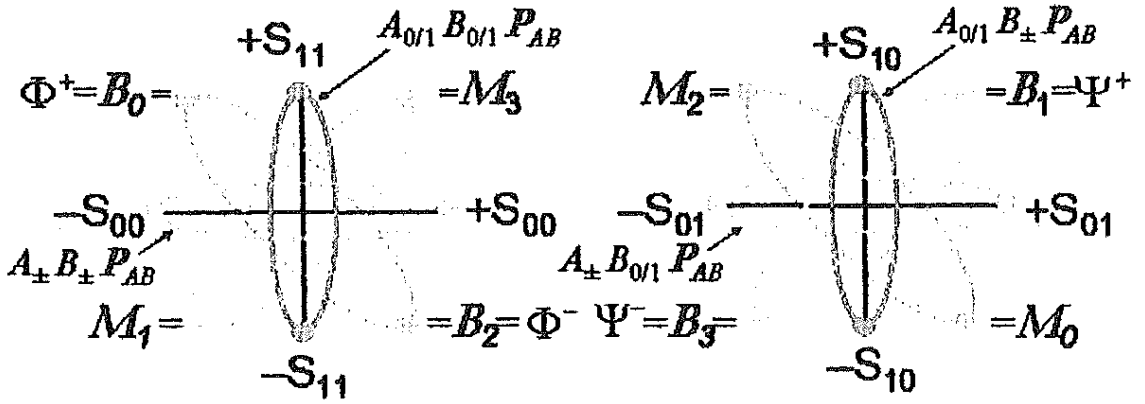


Figure 8. Summary of Bell and Magic States

4.2. *Conditional Operators CNOT and CSPIN*

The only logic-like operator for one qubit is inversion due to phase spinning. The new operators possible for two qubits are the so-called *conditional operators* (similar to the familiar if-then-else clauses) because one qubit acts as a *control qubit* forming a conditional gating state for the operator action on the other *data qubit*. Three or more qubits are required before conventional logic operations can be performed using fully reversible logic gates such as the Toffoli and Fredkin gates.

The conditional form of inversion is called the *control-not operator* (*CNOT*) and the conditional spinor is called the *control-spin*[i] operator (*CSPIN*). Both the *CNOT* and *CSPIN* operators are expressed as multivector operators that are applied using the geometric product. Conditional operators have the general behavior that if the state of a control qubit A is in state A_1 then the operation is performed on data qubit B. Alternately if qubit A is in state A_0 then the operation is *not performed* on qubit B. The *CNOT* operator performs a conditional inversion of the data qubit, while leaving the control qubit unchanged.

Conditional operators are conceptually tricky with regard to quantum computing for the following reasons. First, it is easy to assume, based on classical computing ideas, that in order to "know" the state of the control qubit, it must be measured, which is problematic, if measurement erases information. Second, therefore the conditionality must occur by applying specific operators only to specific states. This is also problematic since the states are thoroughly mixed via entanglement, and it is hard to separate out just the ones you want. Third, geometric products of multivectors are *unconditional* since each n-vector element is jointly affected by every n-vector in the operator. The results achieved so far for one qubit are due to the natural *unconditional* behavior of geometric products, spinors, and destructive interference.

An example of a conditional operator for one qubit is the *reverse*[ii] operator, denoted as $\tilde{A}$. As the name suggests, this operator simply reverses the order of the vectors in an n-vector A, *but* this is not related to the concept of reversibility. If the vectors are then placed back in the standard

[i] Control-spin is usually called a control-Hadamard gate in the literature.

[ii] Reverse is identical to Hermitian adjoint $\tilde{A}=A^{\dagger}$ used in matrices. If $\tilde{A}=A$ then A is self-adjoint

vector order, then *dependent on the overall grade* of the particular n-vector, the coefficient will *conditionally* either remain the same or complement its orientation due to anticommutative operand swaps. The reverse of a multivector is the reverse of each graded element separately, where scalars and vectors are unaffected. Here are some examples.

$$\begin{gathered}\text{reverse}(\pm 1) = \pm 1 \text{ and } \text{reverse}(\mathbf{a}) = \mathbf{a} \\ \text{reverse}(\mathbf{a\,b}) = \mathbf{b\,a} = -\mathbf{a\,b} \\ \text{reverse}(\mathbf{a\,b\,c}) = \mathbf{c\,b\,a} = -\mathbf{a\,b\,c} \\ \text{reverse}(\mathbf{a\,b\,c\,d}) = \mathbf{d\,c\,b\,a} = +\mathbf{a\,b\,c\,d} \\ \text{reverse}(\mathbf{a\,b\,c\,d\,e}) = \mathbf{e\,d\,c\,b\,a} = +\mathbf{a\,b\,c\,d\,e} \\ \text{reverse}(\mathbf{a\,b\,c\,d\,e\,f}) = \mathbf{f\,e\,d\,c\,b\,a} = -\mathbf{a\,b\,c\,d\,e\,f} \end{gathered} \tag{20}$$

Through use of the reverse operator and the operator $A_0 = (\mathbf{a0} - \mathbf{a1})$, a single qubit A can be reversibly encoded into the even-grade plane to represent a *complex number* ($A_0\,A_0 = -1$, $A_1\,A_0 = +1$, $A_-\,A_0 = -\mathbb{S}_A$, $A_+\,A_0 = \mathbb{S}_A$). The operator equivalent to the requisite *complex conjugate* can then be performed using the *reverse* operator to invert conditionally only the sign of the imaginary (or bivector) portion. This result is then converted back into the standard qubit states using the operator $A_1 = (-\mathbf{a0} + \mathbf{a1})$. This sequence of steps $A' = \text{reverse}(A\,A_0)\,A_1$ conditionally inverts only the superposition states $A_\pm$ and topologically represents a reflection of the states off one of the axis, but *cannot be realized* by using only the unconditional geometric product. The main point of this discussion is that in general, writing conditional operators in a reversible linear representation is not straightforward and requires specialized state preparation and operators (e.g. conjugation) other than geometric products. In spite of this general restriction, it is possible to realize *CNOT* and *CSPIN* as multivector operators.

The earlier point regarding *knowing* the state of the control qubit is the inspiration behind the *CNOT* operator. As shown above for the complex number representation of a qubit, it is possible to encode a qubit in the even-grade plane using the operator $A_0 = (\mathbf{a0} - \mathbf{a1})$. The classical states $A_{0/1}$ are mapped to ± 1 respectively (an invariant) and the superposed states $A_\pm$ are mapped to $\pm\mathbb{S}_A$ (a random value). So the result of using *any state as its own operator* is like making a *reversible encoding* without breaking the symmetry of the qubit. This insight is the key to understanding that the control-not operator for control qubit A is $CNOT_{AB} = A_0$. Here are the results of entangling two qubits with the application of the *CNOT* operator.

$$\begin{gathered} A_0\,B\,CNOT_{AB} = (+1)\,B = +B \Rightarrow \text{leave data qubit} \\ A_1\,B\,CNOT_{AB} = (-1)\,B = -B \Rightarrow \text{invert data qubit} \\ A_-\,B\,CNOT_{AB} = (+\mathbb{S}_A)\,B = \mathbb{S}_A(+B) \Rightarrow \text{leave data qubit} \\ A_+\,B\,CNOT_{AB} = (-\mathbb{S}_A)\,B = \mathbb{S}_A(-B) \Rightarrow \text{invert data qubit} \end{gathered} \tag{21}$$

As expected, the *CNOT* operator maps the control qubit to the other encoding, but the *right* multiplication of the operator causes the sign to become inverted due to the non-commutative operation $B\,A_0 = -A_0\,B$. The overall effect is to invert B depending on the state of A. It is useful to think that this reversible operator *reassigns* the information in qubit A to the sign of qubit B (remember $A_0\,B_1 = A_1\,B_0$). So qubit A now contains the state +1, which means *A was classically encoded* and $+\mathbb{S}_A$ means *A was encoded as a superposition.* A control-not gate is intended to be defined only for classical control states, so the result containing the spinor $\mathbb{S}_A$ is correct. The

same analysis derives the operator when the roles are swapped for the data and control qubits. Another way to think of this is that A_1 and B define a simultaneous constraint.

This result is not exactly the conventional definition of the control-not operator since the encoding of the control qubit is modified. This can be remedied if another qubit A' is initialized to the same state $A' = A$, then the result is that the new qubit B includes a duplicate of the entangled information from A, and the qubit A is left intact and untouched. The duplicate must be created in parallel since copying or cloning a qubit requires a measurement. This restriction is called the *no-cloning theorem* of quantum information.

$$A\,A'\,B\,CNOT_{A'B} = A\,(\mp B) = \mp A\,B \tag{22}$$

Since $(S_A)^2 = (spinor)^2 = NOT$ the inspiration occurred to solve for $(CSPIN)^2 = CNOT$, and the result is $CSPIN = \sqrt{CNOT} = -1 + A_0$ (and its other root, and inverse of $+1 + A_1$). This operator has the same concurrent structure as the Pauli spin operator, except with the concurrent operators being the inversion and reversible encoding. Since $CSPIN = \sqrt[4]{-1}$ it indicates a 45 degree rotation. Interestingly, the Bell operators have this exact same structure where $(\mathcal{B})^2 = I^-$, and $\sqrt{\mathcal{B}} = (\mathcal{B})^2 + \mathcal{B} = I^- + \mathcal{B}$ and this structural similarity of equations is most likely a meaningful coincidence. The results of the *CSPIN* operator in Eq. (23) and Table 9 are interesting because they show the need for a *mixed*-grade multivector to encode the phase information.

$$\begin{aligned} A_0\,B_0\,CSPIN_{AB} &= B_0 - A_0\,B_0 = \mathbf{(b0 - b1) - a0\,b0 + a0\,b1 + a1\,b0 - a1\,b1} \\ A_-\,B_0\,CSPIN_{AB} &= S_A\,B_0 + A_+\,B_0 = \mathbf{a0\,a1\,(b0 - b1) + a0\,b0 - a0\,b1 + a1\,b0 - a1\,b1} \end{aligned} \tag{23}$$

For classical states of the control qubit A, Table 9 shows that the overall multivector orientation inverts depending on the control qubit state. The superposition states are also encoded, yet of the 16 possible rows only 6 rows are valid at once. The valid rows indicate what the valid states are and represent a simultaneous constraint system where the operators conditionally change the overall row states that are non-zero. This is clearly evident by the conditional validity of row-states R_5, R_6, R_9 and R_{10} in Table 9.

Table 9. Valid rows for $A\,B\,CSPIN_{AB}$

Row$_k$	Combinations				Active States	$A\,B\,CSPIN_{AB} = -A\,B + B_{0/1}$			
	a0	a1	b0	b1		$A_0B(A_0-1)$	$A_1B(A_0-1)$	$A_0B(A_0+1)$	$A_1B(A_0+1)$
R_1	–	–	–	+	$A_-\&B_1$	+	–	+	–
R_2	–	–	+	–	$A_-\&B_0$	–	+	–	+
R_5	–	+	–	+	$A_1\&B_1$	0	0	– = b0	+ = b1
R_6	–	+	+	–	$A_1\&B_0$	0	0	+ = b0	– = b1
R_9	+	–	–	+	$A_0\&B_1$	– = b0	+ = b1	0	0
R_{10}	+	–	+	–	$A_0\&B_0$	+ = b0	– = b1	0	0
R_{13}	+	+	–	+	$A_+\&B_1$	+	–	+	–
R_{14}	+	+	+	–	$A_+\&B_0$	–	+	–	+

This concludes the new operators for Q_2.

5. *Toffoli Operator is Concurrent CNOT*

The same process for the control-not gate can be expanded to Q_3 in order to include two control qubits A, B and a data qubit D. The resulting *control-control-not gate* is called the Toffoli operator and only inverts qubit D when the control qubits are *both active* (denoted by the subscript 1) in states A_1 and B_1. The individual cases of single control-nots are first expressed to correctly account for the anticommutative operand swaps. The control qubits are indicated by the small subscript c, since it is not always the first one listed in an expression.

$$A\ B_c\ D\ CNOT_{BD} = A\ B_c\ D\ (B_0) = A\ B_c\ B_1\ D = \pm A\ D \quad \text{(one operand swap)}$$
$$A_c\ B\ D\ CNOT_{AD} = A_c\ B\ D\ (A_1) = A_c\ A_1\ B\ D = \pm B\ D \quad \text{(two operand swaps)}$$

Now the *Toffoli Operator is* $TOF_{ABD} = CNOT_{AD} + CNOT_{BD} = A_1 + B_0 = (-\mathbf{a0} + \mathbf{a1} + \mathbf{b0} - \mathbf{b1})$ and is reversible because $(TOF)^2 = +1$. This simple grade-1 multivector operator and grade-2 multivector outcome is a direct result of applying the concurrency interpretation of addition as discovered for the Bell operator. Here is the general Toffoli gate formula:

$$A_c\ B_c\ D\ (TOF_{ABD}) = A_c\ B_c\ D\ (A_1 + B_0) = \pm B\ D \pm A\ D \tag{24}$$

An particular case of Eq. (24) is now required in order to compute the valid rows in Table 10:

$$A_0\ B_0\ D_0\ (TOF_{ABD}) = +\ \mathbf{a0\ d0} - \mathbf{a0\ d1} - \mathbf{a1\ d0} + \mathbf{a1\ d1} + \mathbf{b0\ d0} - \mathbf{b0\ d1} - \mathbf{b1\ d0} + \mathbf{b1\ d1}$$
$$= [00000{+}{-}0\ 0{-}{+}00000\ 0{+}{-}00{-}{+}0\ 00000{+}{-}0\ 0{-}{+}00000\ 0{+}{-}00{-}{+}0\ 00000{+}{-}0\ 0{-}{+}00000] \tag{25}$$

Table 10. Valid row states for $A_0\ B_0\ D_0\ (TOF_{ABD})$ in Q_3

$\mathbf{Row_k}$	State Combinations						Active States	$A_0\ B_0\ D_0$ (TOF_{ABD})	
	a0	**a1**	**b0**	**b1**	**d0**	**d1**			
R_{21}	–	+	–	+	–	+	$A_1\ B_1\ \&\ D_1$	–	Inverted
R_{22}	–	+	–	+	+	–	$A_1\ B_1\ \&\ D_0$	+	
R_{41}	+	–	+	–	–	+	$A_0\ B_0\ \&\ D_1$	+	Identity
R_{42}	+	–	+	–	+	–	$A_0\ B_0\ \&\ D_0$	–	
8 rows	$A_{classical}$		$B_{superpose}$		$D_{classical}$		$A_c\ B_s\ \&\ D_c$	±	Mixed states
8 rows	$A_{superpose}$		$B_{classical}$		$D_{classical}$		$A_s\ B_c\ \&\ D_c$	±	
44 rows	*All conditions not listed above*						*none*	0	Invalid

Rows 21-22 in Table 10 represent the valid states where both control lines are *active* high and the output orientation is inverted compared to qubit D. Rows 41-42 represent the valid states when no inversion occurs, so the output orientation matches qubit D. Since the Toffoli gate $TOF_{ABD} = (-\mathbf{a0} + \mathbf{a1} + \mathbf{b0} - \mathbf{b1})$, it is clear why three qubits in Q_3 are necessary to express this operator. There are four variants of this operator, $A_0 + B_0$, $A_1 + B_0$, $A_1 + B_0$, and $A_1 + B_1$, depending on the desired Boolean condition.

Notice that no other row states are valid when both controls have classical states! This is important because, due to the overall symmetry in geometric algebra, designing arbitrary multiplicative operators is difficult, so in essence *operators are discovered*, not designed. This

problem is akin to building a ship in a bottle, where the quantum state is analogous to a very *high-dimensional bottle* and only tools (or operators) that fit through the neck of the bottle (combinations of single qubit operators) are allowed. It is possible to design an arbitrary state because the row states are linearly independent (given any vector notation can uniquely convert to the algebraic notation and vice versa). Some states can only be created via addition rather than with a multiplicative operator starting from a valid entangled qubit state.

6. Conclusions

The wealth of quantum computing concepts described here, using only addition and geometric products, is possible because geometric algebra naturally and implicitly captures the topological informational distinctions and constraints needed to represent qubits, ebits and familiar operators. This is the interpretation of the co-occurrence of two vectors appears to dominate. Due to the power of geometric algebra to represent classical mechanics, gravitational contraction and quantum mechanics, it is called "a unified language for physics and engineering" [5]. This work extends that domain to include quantum information and quantum computation with straightforward, well-developed [4] and – most importantly – easily interpreted mathematics. This work presents a qubit algebra and as well demonstrates a linearly independent, dual, vector notation that is useful because it combines the topologically smallest elements in the algebra.

It is interesting to see how unfamiliar but transparently meaningful algebraic rules emerge directly from the choice of symmetric binary values +1 and –1 and the mapping of co-occurrence and co-exclusion to addition and the geometric product, i.e. $\mathbf{a}\,\mathbf{b} = -\mathbf{b}\,\mathbf{a}$ and $\mathbf{a}\,\mathbf{a} = 1$. This symmetry then impacts the symmetry of the addition and multiplication operators, i.e. $1/\mathbf{a} = \mathbf{a}$, $2\mathbf{a} = \mathbf{a} + \mathbf{a} = -\mathbf{a} = \mathbf{a}/2$ and enables sparse invariants. This symmetry is reinforced because qubits are the sum of *two* vectors, which results in many counts being a power of 2. As a result, the additive and multiplicative inverses become interchangeable as $A_0 = -A_1 = 1/A_1$, *but also sequential and concurrency ideas herewith intersect,* e.g. $R_k\,R_k = R_k + R_k = P_k$. One should remember that the mathematics describing quantum mechanics is algebraically closed, and so is equivalent to bouncing a light beam around inside a hollow mirrored sphere.

Quantum computing works because it relies on the intrinsically high-dimensional infrastructure of the quantum universe. John Wheeler's paper "It from Bit" [13] stipulates that everything classical, including energy, matter, spacetime and even empty space, emerges from this bit soup (also called quantum ether or quantum foam) because the universe started as a "bit bang" [6,12]. Our geometric algebra approach algebraically and consistently describes topological quantum information forms as a massless high-dimensional topology and true concurrency without focusing on how it is projected into any of the classical properties of space, time or energy. This approach is consistent with extant quantum gravity theories treating the information mechanics of black holes (or bit buckets) [14].

It is possible to make better decisions, *to be smarter,* with high-dimensional spaces [15] because more states can participate simultaneously in a decision, due to a higher locality metric and true concurrency. Quantum metrics and phenomena are not possible in computation restricted to classical spacetime. Spacetime itself limits the computational density by segregating [16] the required information locality and concurrency. This alone should motivate engineers and

programmers to want to understand quantum computing: because it allows computers to *cheat* by computing outside the limiting spacetime box that occurs when representing bits classically. Because of the unusual and counterintuitive nature of quantum information, encouraging engineers and programmers to ascend the quantum computing learning curve will lead to an appreciation of the fundamental role of information in the quantum computing universe and might lead to general purpose quantum computers.

Acknowledgements
Special thanks go to my UTD Ph.D. committee chairmen and members, especially Mike Manthey for teaching me about co-occurrence and co-exclusion principles in the GA context. Also, I would like to thank Katrina Riehl, who helped outline the sequence of quantum computing concepts that would be useful (and also avoided) for a novice computer professional.

References

1. P. Shor (1994), "Algorithms for Quantum Computation: Discrete Logarithms and Factoring", In *Proceedings of 35th Annual Symposium on the Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, CA, page 124.
2. F. Mattern (1999), "Quantum Computing Introduction", Paper found on his website http://www.rommel.stw.uni-erlangen.de/~frank/informatik/QuantumComputing.pdf.
3. D. Hestenes (1999), *New Foundations for Classical Mechanics* (Second Edition), Kluwer Academic Press.
4. D. Matzke (May 2002), "Quantum Computation using Geometric Algebra", University of Texas at Dallas, Ph.D. dissertation in the Department of Electrical Engineering. See http://www.photec.org.
5. J. Lasenby, A.N. Lasenby and C.J.L. Doran (2000), "A unified mathematical language for physics and engineering in the 21st century", *Phil. Trans. R. Soc. Lond*, A358, pp. 21-39.
6. M. Manthey (Sept 1998), "A Combinatorial Bit Bang Leading to Quaternions". See paper number 9809033 on LANL Eprints server at http://eprints.lanl.gov.
7. R. Landauer (1991), "Information is Physical", *Physics Today*, Vol. 44, pp. 23-29.
8. C. Bennett (1973), "Logical Reversibility of Computation", *IBM Journal of Research and Development*. Vol. 17, pp. 525-532.
9. C. D. Cantrell (2000), *Modern Mathematical Methods for Physicists and Engineers*, Cambridge University Press.
10. C. J. L. Doran, (2000) Handouts for course "Physical Applications of Geometric Algebra", See http://www.mrao.cam.ac.uk/~clifford/ptIIIcourse/, Handout for lecture 4 on Geometric Algebra and Quantum Mechanics, Section 2 on "Spinors and Multivectors".
11. J. Bell (1964), "On the Einstein-Podolsky-Rosen Paradox", *Physics*, Vol. 1, pp. 195-200.
12. D. Matzke (1996), "Information is Protophysical", *Proceedings of the Workshop on Physics and Computation, PhysComp96*, New England Complex System Institute.
13. J. Wheeler (1989), "It From Bit", *Proceedings of the 3rd International Symposium on Foundations of Quantum Mechanics*, Tokyo.
14. M. Schiffer (1993), "The interplay between Gravitation and Information Theory", *Proc. of the Workshop on Physics and Computation, PhysComp92*, IEEE Computer Society Press.
15. P. Kanerva (1988), *Sparse Distributed Memory*, MIT Press.
16. D Matzke (September 1997), "Will Physical Scalability Sabotage Performance Gains?", *Computer Magazine*, Vol. 30, No. 9, pp 37-39.

Non-commutative Calculus and Discrete Physics

Louis H. Kauffman
Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
851 South Morgan Street
Chicago, IL, 60607-7045

March 10, 2003

1 Introduction

This paper is an expanded version of [28] and [30] where there is presented an introduction to a point of view for discrete foundations of physics. In taking a discrete stance, we find that the initial expression of physical observation naturally occurs in a context of non-commutative algebra. In this way a formalism similar to quantum mechanics occurs first, but not necessarily with the usual interpretations. By following this line we show how the outlines of the well-known forms of physical theory arise first in non-commutative form. The exact relation of commutative and non-commutative theories raises a host of problems.

The starting point for this investigation is the representation of calculus in a non-commutative framework. In such a framework derivatives are represented by commutators, or more generally by products that satisfy the Jacobi identity and the Leibniz rule. If we take commutators $[A, B] = AB - BA$ in an abstract algebra and define $DA = [A, J]$ for a fixed element J, then D acts like a derivative in the sense that $D(AB) = D(A)B + AD(B)$ (the Leibniz rule). As soon as we have calculus in such a framework, concepts of geometry are immediately available. For example, if we have two derivatives $\nabla_J A = [A, J]$ and $\nabla_K A = [A, K]$, then we can consider the commutator of

these derivatives $[\nabla_J, \nabla_K]A = \nabla_J \nabla_K A - \nabla_K \nabla_J A = [[J, K], A]$. The non-commutation of derivations corresponds to curvature in geometry, and indeed we shall see that the earliest emergence of curvature in this context is the formal analog of the curvature of a gauge connection!

For multivariable calculus we need variables $X_1, X_2, \cdots, X_n$ and elements $P_1, P_2, ..., P_n$ such that $\partial_i A = \partial A / \partial X_i = [A, P_i]$. For a simplest representation we shall assume the the X_i commute with one another, and that the P_j commute with one another. Since we want $\partial_i X_j = \delta_{ij}$ (the Kronecker delta δ_{ij} is equal to one if i and j are equal and is zero otherwise), we must have the commutator equation $[X_i, P_j] = \delta_{ij}$. Thus multivariable calculus in this non-commutative representation demands the commutation relations

$$[X_i, X_j] = 0$$

$$[P_i, P_j] = 0$$

$$[X_i, P_j] = \delta_{ij}$$

These equations are the "flat background" for our non-commutative calculus. The reader will note that this flat background has the same pattern of commutation relations as a bare form of quantum mechanics when the X variables are interpreted as position and the P variables are interpreted as momenta. In a certain sense this means that our considerations start in the quantum domain. Note that flat is a correct adjective, since the derivatives ∂_i all commute with one another.

Let A_i be a collection of elements of this algebra. Define "covariant derivatives" with $\Lambda_i = P_i - A_i$ by the formula

$$\nabla_i Z = [Z, \Lambda_i] = \partial_i Z - [Z, A_i].$$

Computing the curvature, one finds

$$[\nabla_i, \nabla_j] Z = [[\Lambda_i, \Lambda_j], Z]$$

and

$$[\Lambda_i, \Lambda_j] = \partial_i A_j - \partial_j A_i + [A_i, A_j].$$

The reader will recognize this last expression as the formula for the curvature of a gauge connection.

In interfacing this formalism with physics we adopt the coupling equation

$$dX_i/dt = \Lambda_i = P_i - A_i.$$

The reader will recognise this as the minimal coupling postulate in the context of Poisson brackets. Here we take it in the context of commutators or Poisson brackets, or a more general product satisfying the Jacobi identity and the Leibniz rule as described above. One retrieves the physics of a gauge field in this formalism. This is the essence of the pattern behind the Feynman-Dyson derivation of electromagnetism from commutation relations [24, 27], and its import is more general. Because the brackets can be interpreted as commutators or as Poisson brackets with special structure, the formalism can be seen in a multiplicity of contexts. Deeper relationships with curvature and metric are related to this shifting of contexts as are relationships with quantum mechanics where the quantum formalism is obtained by the Dirac prescription of replacing Poisson brackets by commutators. We will discuss these issues in Section 5 of this paper. The organization of the paper is as follows.

Section 2 of this paper we discuss the properties of the non-commutative discrete calculus that underlies our work. Here we begin with the consideration of a temporal operator J with the property that $YJ = JY'$ for a "time series" $X, X', X'', \cdots$. Thus $XJ = JX', X'J = JX'', \cdots$. This formalism for time series gives rise to the time derivative $DA = [A, J] = AJ - JA = JA' - JA = J(A' - A)$, a commutator representing a discrete derivative. Note that DA satisfies the Leibniz rule, a privilege not shared by the usual commutative discrete derivative. This section discusses the discrete ordered calculus (DOC) that arises from this idea and applies these ideas to a number of situations. In particular, we consider the one variable case of the commutator equation $[X, DX] = Jk$ and show that it leads to a Brownian walk, and that if we take the size of the time step into account, then the diffusion constant for a Brownian process arises naturally as $k/2$. We compare this with the usual derivation of the diffusion constant and the diffusion differential equation. We then compare this situation with the one dimensional Schrödinger equation, modeling it in relation to a diffusion process with complex amplitudes. In this viewpoint one sees that the step length of the diffusion process is the Compton wavelength associated with the mass for the particle, and the time is the Compton time. For the Planck mass this gives a step equal to the Planck length and a time interval equal to the Planck

time. We speculate on the relationship of this result to joint work with Pierre Noyes and others [37]. We consider other time series that can be regarded as solutions to this Heisenberg relation, the problem of using more variables and a model that is related to a discrete version of the Feynman-Dyson derivation of electromagnetic formalism.

Section 3 examines the consequences for a particle whose position - momentum commutator is equated to a metric field. Here we see how the Levi-Civita connection (and implicitly differential geometric structure) comes naturally from the non-commutative calculus. This is a very general result and in section 4 we discuss it in a more axiomatic context as described in this introduction. This section discusses the intimate relationship between that Levi-Civita connection and the Jacobi and Leibniz identities that is revealed by our non-commutative calculus. In section 5 our stance leads to an inversion of the usual Dirac maxim "replace Poisson brackets with commutators". If we replace commutators with Poisson brackets that obey a Leibniz rule satisfied by the commutators, then the dynamical variables will obey Hamilton's equations. Thus we can take Hamilton's equations as a classicization of our theory. Among other things, this point of view explains the appearance of the Levi-Civita connection in the abstract formalism. Interpreting with Poisson brackets, we obtain a new proof (via Jacobi identity) of the classical result that a Newtonian particle moving in generalized coordinates according to Lagrange's equations describes a geodesic in the Levi-Civita connection. Section 6 discusses the relationship of the discrete ordered calculus with q-deformations and quantum groups. We show that in a quantum group with a special grouplike element representing the square of the antipode, there is a representation of the discrete ordered calculus. In this calculus on a quantum group the square of the antipode can represent one tick of the clock. Then follows section 7 on networks and discrete spacetime. This section is an exposition of ideas related to spin networks and topological quantum field theory. As an early example we discuss the discretization of the Dirac equation in $1+1$ dimensional spacetime. It is our speculation that the approaches to discrete physics inherent in discrete calculus and in topological field theory are deeply interrelated. At the end of this section we outline this relationship in the case of a model for quantum gravity due to Louis Crane. Section 8 is an appendix on the iterant approach to matrix algebra. We include this appendix to show how one can conceptualize matrix algebra from point of view of the discrete. Section 9 is a philosophical appendix discussing the

nature of foundations in mathematics and in physics.

Remark. The following references in relation to non-commutative calculus are useful in comparing with our approach [7, 10, 13, 35]. Much of the present work is the fruit of a long series of discussions with Pierre Noyes, and we will be preparing collaborative papers on it. The present paper is a summary for the proceedings of the ANPA Conference held in Cambridge, England in the summer of 2002. I particularly thank Eddie Oshins for pointing out the relevance of minimal coupling. The paper [34] also works with minimal coupling for the Feynman-Dyson derivation. The first remark about the minimal coupling occurs in the original paper by Dyson [4], in the context of Poisson brackets. The paper [17] is worth reading as a companion to Dyson. In the present paper we generalize the minimal coupling to contexts including both commutators and Poisson brackets. The reader can see the full generality of our approach by first reading this introduction and then going directly to sections 4 and 5. It is the purpose of this paper to indicate how non-commutative calculus can be used in foundations.

Acknowledgement. Most of this effort was sponsored by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory, Air Force Materiel Command, USAF, under agreement F30602-01-2-05022. Some of this effort was also sponsored by the National Institute for Standards and Technology (NIST). The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright annotations thereon. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Defense Advanced Research Projects Agency, the Air Force Research Laboratory, or the U.S. Government. (Copyright 2003.) It gives the author pleasure to thank Pierre Noyes, Clive Kilmister, Ted Bastin, Tony Deakin, Eddie Oshins, Basil Hiley, Keith Bowden, Arleta Giffor, Ashok Gangadean, Lynnclaire Dennis, Louis Licht and Sam Lomonaco for many conversations during the course of this work, and the Theory Group of the Stanford Linear Accelerator Laboratory for hospitality during the preparation of parts of the present paper.

2 Discrete Ordered Calculus

In this section we recall the construction of an ordered version of the calculus of finite differences *DOC* [24], [28]. In this calculus the Leibniz rule is satisfied, and so the calculus can be used in a variety of applications.

In the abstract framework of this calculus, there are variables X, each of which connotes a time series

$$X, X', X'', ...$$

Discrete unit time steps are indicated by the primes appended to the X. A general point in the time series at time t will be denoted by X^t. By convention let the time step between successive points in the series be equal to 1 :

$$\Delta t = 1.$$

Then we can define the velocity at time t by the formula:

$$v(t) = X^{t+1} - X^t.$$

More generally, if X denotes position at a given time, then $X' - X$ denotes the velocity *at that time*, where the phrase "at that time" must involve the next time as well. In a discrete context there is no notion of instantaneous velocity.

Measure position, and you find X. Then measure velocity, and you get $X' - X$. Now measure position, and you get X' because the time has shifted to the next time in order to allow the velocity measurement. In order to measure velocity the position is necessarily shifted to its value at the next time step. In this sense, *position and velocity measurements cannot commute in a discrete framework.* This is the key physical idea that motivates our constructions. It was this idea, told to the author by Pierre Noyes, that led to our papers and particularly to [24].

The simplest interpretation of the variable X is that the time series values are numerical values, commuting with one another and with any operators that might be present in the associated mathematics or physics. In fact, we will often deal with situations where the X and the elements of the time series are in fact operators, not necessarily commuting with one another. At

the very least we will construct an algebra that mirrors the discrete non-commutativity of the operations of position and velocity measurement.

Our project is to take this basic noncommutativity at face value and follow out its consequences. To this end we will formulate a calculus of finite differences that takes the order of observations into account. This formalization is explained below.

To see most clearly the non-commutativity that is at the base of our considerations, let J denote the operation of shifting time by one increment. Thus we can envisage an algebra of operations that consists in commands like JX (measure X, then tick the clock). *Note that we will agree to take the sequence of operations from right to left.* Let $|JX|$ denote the "spatial evaluation" of this sequence of operations, obtained in general by performing all the instructions and then evaluating the spatial position. Thus

$$|JX| = X$$

while

$$|XJ| = X'$$

since when the clock ticks, the position shifts to the position at the next time. We see therefore, that $XJ \neq JX$. This is the first instance of non-commutativity in the physics of discrete space and time. From the point of view of spatial evaluation it is most convenient to declare the equation

$$XJ = JX'$$

since these two expressions have identical spatial evaluations.

We can then define the DOC derivative by the equation

$$DX = [X, J] = XJ - JX = JX' - JX = J(X' - X) = JdX$$

where dX denotes the classical discrete derivative with unit time step. The key point about the DOC derivative is that it is a commutator, and consequently satisfies the Leibniz rule

$$D(XY) = D(X)Y + XD(Y).$$

This makes it possible to do discrete calculus in a way that is formally similar to classical calculus. We will repeat this structure more slowly now, first recalling the properties of classical discrete derivatives.

We begin by recalling the usual derivative in the calculus of finite differences, generalised to a (possibly) non-commutative context.

Definition. Let

$$dX = X' - X$$

define the finite difference derivative of a variable X whose successive values in discrete time are

$$X, X', X'', ...$$

This dX is a classical derivative in the calculus of finite differences. It is still defined even if the quantities elements of the time series are in a non-commutative algebra. We shall assume that the values of the time series are in a possibly non-commutative ring R with unit. (Thus the values could be real numbers, complex numbers, matrices, linear operators on a Hilbert space, or elements of an appropriate abstract algebra.) This means that for every element A of the ring R there is a well-defined successor element A', the next term in the time series. It is convenient to assume that the ring itself has this temporal structure. In practice, one is concerned with a particular time series and not the structure of the entire ring. Moreover, we shall assume that the next-time operator distributes over both addition and multiplication in the sense that

$$(A + B)' = A' + B'$$

and

$$(AB)' = A'B'.$$

An element c of the ring R is said to be a *constant* if $c' = c$.

Lemma 1.

$$d(XY) = X'd(Y) + d(X)Y.$$

Proof.

$$\begin{aligned} d(XY) &= X'Y' - XY \\ &= X'Y' - X'Y + X'Y - XY \\ &= X'(Y' - Y) + (X' - X)Y \\ &= X'd(Y) + d(X)Y. \end{aligned}$$

This formula is *different* from the usual formula in Newtonian calculus by the time shift of X to X' in the first term. We now correct this discrepancy in the calculus of finite differences by taking a *new* derivative D as an *instruction to shift the time to the left of the operator* D. That is, we take $XD(Y)$ quite literally as an instruction to *first find* dY *and then find the value of* X. In order to find dY the clock must advance one notch. Therefore X has advanced to X' and we have that the evaluation of $XD(Y)$ is

$$X'(Y' - Y).$$

In order to keep track of this non-commutative time-shifting, we will write

$$DX = J(X' - X)$$

where the element J is a special time-shift operator satisfying

$$ZJ = JZ'$$

for any Z in the ring R. The time-shifter, J, acts to automatically evaluate expressions in the resulting non-commutative calculus of finite differences. We call this calculus DOC (for discrete ordered calculus). Note that J formalizes the operational ordering inherent in our initial discussion of velocity and position measurements. An operator containing J causes a time shift in the variables or operators to the left of J in the sequence order.

Formally, we extend the ring of values R (see the definition of d above) by adding a new symbol J with the property that $AJ = JA'$ for every A in R. It is assumed that the extended ring R is associative and satisfies the distributive law so that $J(A + B) = JA + JB$ and $J(AB) = (JA)B$ for all A and B in the ring. We also assume that J itself is a constant in the sense that $J' = J$.

The key result in DOC is the following adjusted difference formula:

Lemma 2.

$$D(XY) = XD(Y) + D(Y)X.$$

Proof.

$$\begin{aligned} & D(XY) \\ & = J(X'Y' - XY) \\ & = J(X'Y' - X'Y + X'Y - XY) \\ & = J(X'(Y' - Y) + (X' - X)Y) \\ & = JX'(Y' - Y) + J(X' - X)Y \\ & = XJ(Y' - Y) + J(X' - X)Y \\ & = XD(Y) + D(X)Y. \end{aligned}$$

The upshot is that *DOC* behaves formally like infinitesimal calculus and can be used as a calculus in this version of discrete physics. In [24] Pierre Noyes and the author use this foundation to build a derivation of a non-commutative version of electromagnetism. Another version of this derivation can be found in [27]. In both cases the derivation is a translation to this context of the well-known Feynman-Dyson derivation of electromagnetic formalism from commutation relations of position and velocity.

Note that the definition of the derivative in *DOC* is actually a commutator:

$$DX = J(X' - X) = JX' - JX = XJ - JX = [X, J].$$

The operator J can be regarded as a discretised time-evolution operator in the Heisenberg formulation of quantum mechanics. In fact we can write formally that

$$X' = J^{-1}XJ$$

since $JX' = XJ$ (assuming for this interpretation that the operator J is invertible). Putting the time variable back into the equation, we get the evolution

$$X^{t+\Delta t} = J^{-1}X^t J.$$

This aspect can be compared to the formalism of Alain Connes' theory of non-commutative geometry [7].

In *DOC*, X and DX have no reason to commute:

$$[X, DX] = XJ(X' - X) - J(X' - X)X = J(X'(X' - X) - (X' - X)X)$$

Hence

$$[X, DX] = J(X'X' - 2X'X + XX).$$

This is non-zero even in the case where X and X' commute with one another. Consequently, we can consider physical laws in the form

$$[X_i, DX_j] = g_{ij}$$

where g_{ij} is a function that is suitable to the given application. In [24] we show how the formalism of electromagnetism arises when g^{ij} is δ^{ij}, the Kronecker delta. In [26] we will show how the general case corresponds to a "particle" moving in a non-commutative gauge field coupled with geodesic motion relative to the Levi-Civita connection associated with the g_{ij}. This result can be used to place the work of Tanimura [42] in a discrete context.

It should be emphasized that all physics that we derive in this way is formulated in a context of non-commutative operators and variables. We do not derive electromagnetism, but rather a non-commutative analog. It is not yet clear just what these non-commutative physical theories really mean. Our initial idealisation of measurement is not the only model for measurement that corresponds to actual observations. Certainly the idea that we can measure time in a way that has "steps between the steps of time" is an idealisation. It happens to be an idealisation that fits a model of the universe as a cellular automaton. In a cellular automaton an observation is what an operator of the automaton might be able to do. It is not necessarily what the "inhabitants" of the automaton can perform. Here is the crux of the matter. The inhabitants can have only limited observations of the running of the automaton, due to the fact that they themselves are processes running on the automaton. The theories we build on the basis of *DOC* can be theories *about* the structure of these automata. They will eventually lead to theories of what can be observed by the processes that run on such automata. It is possible that the well known phenomena of quantum mechanics will arise naturally in such a context. These points of view should be compared with [14].

2.1 Brownian Walks and the Diffusion Equation

To return to basics, consider the commutator equation in one space variable X.

$$[X, DX] = Jk$$

for a single variable X. Written out, this equation becomes

$$Jk = [X, J(X'-X)] = XJ(X'-X)-J(X'-X)X = J(X'(X'-X)-(X'-X)X).$$

If k and the elements of the time series $\{X, X', X'', ...\}$ are all commuting scalars then this equation reduces to

$$k = (X' - X)^2.$$

Thus

$$X' = X \pm k^{1/2},$$

a Brownian random walk, is a solution to the simplest one-dimensional commutator equation.

Now lets examine this Brownian walk more closely by quantifying the time step as well as the space step. We take

$$\Delta t = \tau$$

so that

$$DX = J(X' - X)/\tau$$

where it is assumed that τ is a scalar, commuting with all elements of the time series and commuting with the operator J (that is, τ does not change with time). Now examine once again the equation

$$[X, DX] = Jk.$$

Let $|X' - X| = \Delta$. Then, repeating the calculation, we find

$$k = (X' - X)^2/\tau = \Delta^2/\tau.$$

Hence

$$\Delta^2/\tau = k.$$

This tells us that if k is to be constant then there must be a constant relationship between the square of the space interval for the Brownian walk and the size of the time interval. The remarkable point here is that it is just this

constant relationship that is required for a Brownian process to be described by the diffusion equation

$$\partial P(x,t)/\partial t = C\partial^2 P(x,t)/\partial x^2$$

where the diffusion constant C is given by the formula

$$C = \Delta^2/2\tau = k/2.$$

The diffusion constant comes directly from our consideration involving the DOC commutator without any of the usual conceptual apparatus about approximating a differential equation.

To make this comparison, lets recall how the diffusion equation usually arises in discussing Brownian motion. We are given a Brownian process where

$$x(t+\tau) = x(t) \pm \Delta$$

so that the time step is τ and the space step is of absolute value Δ. We regard the probability of left or right steps as equal, so that if $P(x,t)$ denotes the probability that the Brownian particle is at point x at time t then

$$P(x,t+\tau) = P(x-\Delta,t)/2 + P(x+\Delta)/2.$$

From this equation for the probability we can write a difference equation for the partial derivative of the probability with respect to time:

$$[(P(x,t+\tau) - P(x,t))/\tau] = (h^2/2\tau)[(P(x-\Delta,t) - 2P(x,t) + P(x+\Delta))/\Delta^2]$$

The expression in brackets on the right hand side is a discrete approximation to the second partial of $P(x,t)$ with respect to x. Thus if the ratio $C = \Delta^2/2\tau$ remains constant as the space and time intervals approach zero, then this equation goes in the limit to the diffusion equation

$$\partial P(x,t)/\partial t = C\partial^2 P(x,t)/\partial x^2.$$

It is most curious how the diffusion constant comes up in these two contexts. Lets try to think about the comparison between the non-commutative observational starting point and the more standard differential approximation. In the non-commutative context we get Δ^2 from the appearance of the square of the difference of X' and X in the calculation of the commutator of X and DX. In the differential approximation, we get the Δ^2 from the approximation of the second derivative of the probability $P(x,t)$ with respect to x. The concept of probability does not appear in the non-commutative context. Clearly this subject needs more thought.

2.2 Planck's Numbers, Schrödinger's Equation and the Diffusion Equation

First recall the Planck Numbers. $\hbar$ is Planck's constant divided by 2π. c is the speed of light. G is Newton's gravitational constant. The Planck length will be denoted by L, the Planck time by T and the Planck mass by M. Their formulas are

$$M = \sqrt{\hbar c/G}$$
$$L = \hbar/Mc$$
$$T = \hbar/Mc^2.$$

These amounts of mass, length and time have just these dimensions and are constructed from the values of fundamental physical constants. They have roles in physics that point to deeper reasons than the formal for introducing them. Here we shall see how they are related to the Schrödinger equation.

Recall that Schrödinger's equation can be regarded as the diffusion equation with an imaginary diffusion constant. Recall how this works. The Schrödinger equation is

$$i\hbar\partial\psi/\partial t = H\psi$$

where the Hamiltonian H is given by the equation $H = p^2/2m + V$ where $V(x,t)$ is the potential energy and $p = \hbar/i\partial/\partial x$ is the momentum operator. With this we have $p^2/2m = (-\hbar^2/2m)\partial^2/\partial x^2$. Thus with $V(x,t) = 0$, the equation becomes $i\hbar\partial\psi/\partial t = (-\hbar^2/2m)\partial^2\psi/\partial x^2$ which simplifies to

$$\partial\psi/\partial t = (i\hbar/2m)\partial^2\psi/\partial x^2.$$

Thus we have arrived at the form of the diffusion equation with an imaginary constant, and it is possible to make the identification with the diffusion equation by setting

$$\hbar/m = \Delta^2/\tau$$

where Δ denotes a space interval, and τ denotes a time interval as explained in the last section about the Brownian walk. With this we can ask what space interval and time interval will satisfy this relationship with a mass and Planck's constant? *Remarkably, the answer is that this equation is satisfied when m is the Planck mass, Δ is the Planck length and τ is the Planck time!!* For note that

$$L^2/T = (\hbar/Mc)^2/(\hbar/Mc^2) = \hbar/M.$$

I now quote an email comment of Pierre Noyes: "With regard to your DOC derivation of the diffusion equation, and with an imaginary diffusion coefficient, the Schrödinger equation, note that the relation $\hbar/m = L^2/T$ is satisfied for *any* mass m provided we take L = Compton wavelength = $\hbar/mc$ and T =Compton time= $\hbar/mc^2$ — which is simply the time of a step length of this length taken at the velocity of light. I have a vague idea that I heard of this relation when I was a graduate student. In any case I am sure Feynman had it in mind when he used a random walk on the light cone to derive the 1+1 Dirac equation, and counted steps using i! So, in a sense, your DOC derivation of the diffusion equation does connect the Maxwell equations derivation via DOC, to the Dirac equation derivation — which in a vague sense was what I hoped we would be able to do this spring (2002). Of course this general result applies in particular to the Planck mass, which was your first observation. It is intriguing that if the mass scale is m [the Planck mass], then we can use either the Compton wavelength or the Schwarzchild radius at that mass scale as the step length in DOC. This reinforces my conviction (expressed long ago) that elementary particles are small black holes."

The last part of Noyes' remark about the Schwarzchild radius refers to our work [37] explaining Ed Jones' microcosmology. Jones observed that if, for a particle of mass m we set the Schwarzchild radius ($R_S = 2mG/c^2$) equal to the Compton radius ($R_C = \hbar/2mc$), then the resulting mass m is equal to one half the Planck mass!

$$R_S = R_C$$
$$2mG/c^2 = \hbar/2mc$$
$$m = (1/2)\sqrt{\hbar c/G} = M/2$$

This is highly suggestive of limiting conditions on matter ("Plancktonic matter") prior to the Big Bang and leads in this way to specific cosmological predictions. It also gives an intriguing physical meaning to the Planck mass.

What does all this say about the nature of the Schrödinger equation itself? Interpreting it as a diffusion equation with imaginary constant suggests comparing with the DOC equation

$$[X, DX] = JiC$$

for a real constant C. This equation implicates a Brownian process where $X' = X \pm Z$ where $Z^2/\tau = iC$. We can take $Z = \sqrt{i}L$ where L is a real step-length. This gives a Brownian walk in the complex plane with the correct

DOC diffusion constant. However, the relationship of this walk with the Schrödinger equation is less clear because the ψ in that equation is not the probability for the Brownian process. To see a closer relationship we will take a different tack.

Consider a discrete function $\psi(x,t)$ defined (recursively) by the following equation

$$\psi(x,t+\tau) = (i/2)\psi(x-\Delta,t) + (1-i)\psi(x,t) + (i/2)\psi(x+\Delta,t)\,.$$

In other words, we are thinking here of a random "quantum walk" where the amplitude for stepping right or stepping left is proportional to i while the amplitude for not moving at all is proportional to $(1-i)$. It is then easy to see that ψ is a discretization of

$$\partial\psi/\partial t = (i\Delta^2/2\tau)\partial^2\psi/\partial x^2.$$

Just note that ψ satisfies the difference equation

$$(\psi(x,t+\tau)-\psi(x,t))/\tau = (i\Delta^2/2\tau)(\psi(x-\Delta,t)-2\psi(x,t)+\psi(x+\Delta,t))/\Delta^2.$$

This gives a direct interpretation of the solution to the Schrödinger equation as a limit of a sum over generalized Brownian paths with complex amplitudes. We can then reinterpret this in DOC terms by the equation $[X,DX] = J(\Delta^2/\tau)$ or $[X,DX] = 0$, each of these contingencies happening probabilistically. It remains to be seen whether there is further insight to be gained into the Schrödinger equation via this combination of the DOC approach and the stochastic approach.

2.3 DOC Chaos

Along with the simple Brownian motion solution to the one dimensional commutator equation, there is a hierarchy of time series that solve this equation, with periodic and chaotic behaviour. These solutions can be obtained by taking

$$X = J^n Y$$

where Y is a numerical scalar, and taking the commutator equation to be

$$[X,DX] = J^{2n+1}k$$

where k is a scalar. Expanding this equation, we find

$$XJ(X'-X)-J(X'-X)X=J^{2n+1}k$$

$$J^nYJ(J^nY'-J^nY)-J(J^nY'-J^nY)J^nY=J^{2n+1}k$$

$$J^{2n+1}Y^{n+1}(Y'-Y)-J^{2n+1}(Y^{n+1}-Y^n)Y=J^{2n+1}k$$

$$Y^{n+1}(Y'-Y)-(Y^{n+1}-Y^n)Y=k$$

$$Y^{n+1}(Y'-2Y)=k-Y^nY$$

$$Y^{n+1}=(k-Y^nY)/(Y'-2Y).$$

This last equation expresses the time series recursively where Y refers to the value of the series that is n time steps back from Y^n. The first case of this recursion is

$$Y''=(k-Y'Y)/(Y'-2Y).$$

Next case is

$$Y'''=(k-Y''Y)/(Y'-2Y).$$

These recursions depend critically on the value of the parameter k. In the first case one sees periodic oscillations that (for appropriate values of k) destabilize and blow up, alternating between an unbounded phase and a bounded semi-periodic phase. We will investigate these time series in a separate paper.

2.4 More Variables

In the Feynman-Dyson derivation of electromagnetic formalism from commutation relations [24] one uses the relations

$$[X_i,X_j]=0$$

$$[X_i,DX_j]=k\delta_{ij}$$

where k is a scalar. Here we shall use

$$[X_i,X_j]=0$$

$$[X_i,DX_j]=Jk\delta_{ij}$$

as we did in analyzing the one-dimensional case. This allows us to have scalar evolution of the time series, but changes some of the issues in the

Feynman-Dyson derivation. These are in fact handled by the more general formalism that we discuss in the next two sections. Thus we shall aim in this section to see to what extent one can make simple models for this version of the Feynman-Dyson relations. Models of this sort will be another level of approximation to discrete electromagnetism.

Writing out the commutation relation $[X, DX] = Jk$, and not making any assumption that X' commutes with X, we find

$$J^{-1}[X, DX] = X'(X' - X) - (X' - X)X$$
$$= X'(X' - X) - X(X' - X) + X(X' - X) - (X' - X)X$$
$$= (X' - X)^2 + (XX' - X'X) = (X' - X)^2 + [X, X'].$$

Thus the commutation relation $[X, DX] = Jk$ becomes the equation

$$(X' - X)^2 + [X, X'] = k.$$

By a similar calculation, the equation $[X, DY] = 0$ becomes the equation

$$(X' - X)(Y' - Y) + [X, Y'] = 0.$$

These equations are impossible to satisfy simultaneously for $k \neq 0$ if we assume that X and X' commute and that X and Y' commute and that $[Y, DY] = Jk$. For then we would need to solve:

$$(X' - X)^2 = k,$$
$$(Y' - Y)^2 = k,$$
$$(X' - X)(Y' - Y) = 0,$$

with the first two equations implying that $(X - X')$ and $(Y - Y')$ are each non-zero, and the third implying that their product is equal to zero. In other words, the equations below cannot be satisfied if the time series are composed of commuting scalars.

$$[X, DX] = Jk$$
$$[Y, DY] = Jk$$
$$[X, Y] = 0$$

In order to make such models we shall have to introduce non-commutativity into the time series themselves. In a certain sense this is analoguous to the introduction of non-commutative algebra in the Dirac equation in $3 + 1$ dimensions, and to the introduction of non-commutative fields in gauge theory.

Here is an example of such a model.
Return to the equations

$$(X' - X)^2 + [X, X'] = k.$$

$$(X' - X)(Y' - Y) + [X, Y'] = 0$$

expressing the behaviour for two distinct variables X and Y. If $[X, X'] = 0$, then we have $(X' - X)^2 = k$ so that

$$X' = X \pm \sqrt{k}.$$

In order for the second equation to be satisfied, we need that

$$[X, Y'] = \pm k$$

where the ambiguity of sign is linked with the varying signs in the temporal behaviour of X and Y. We will make the sign more precise in a moment, but the radical part of this suggestion is that for two distinct spatial variables X and Y, there will be a commutation relation between one and a time shift of the other.

If the space variables are labeled X_i, then we can write

$$X_i^{t+1} = X_i^t + \epsilon_i^t k$$

where ϵ_i^n is plus one or minus one. Thus each space variable performs a walk with the fixed step-length k. We shall write informally

$$X_i' = X_i + \epsilon_i k$$

where it is understood that the epsilon without the superscript connotes the sign change that occurs in this juncture of the process. We then demand the commutation relations

$$[X_i', X_j] = [X_j', X_i] = \epsilon_i \epsilon_j k.$$

Each X_i is a scalar in its own domain, but does not commute with the time shifts of the other directions. We then can have the full set of commutation relations:

$$[X_i', X_j] = [X_j', X_i] = \epsilon_i \epsilon_j k.$$

$$[X_i, X_j] = 0$$

$$[X_i, DX_j] = Jk\delta_{ij}$$

so that the system will satisfy the assumptions supporting the Feynman-Dyson derivation. In this system,the elements of a given time series $X_i, X_i', X_i'', \cdots$ commute with one another. The basic field element in the Feynman-Dyson set up is the magnetic field B defined by the (non-commutative) vector cross product

$$B = (1/k)DX \times DX.$$

Here we have

$$DX_i = J(X_i' - X_i) = J\epsilon_i\sqrt{k}.$$

Thus

$$B = J^2\ \epsilon' \times \epsilon$$

where $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)$ (assuming three spatial coordinates) and ϵ' denotes this vector of signs at the next time step. In this way we see that we can think of each spatial coordinate as providing a long temporal bit string and the three coordinates together give the field in terms of the vector cross product of their temporal cross sections at neighboring instants. It is interesting to compare this model with the color algebra in the following paper by Wene [43].

2.5 Discrete Classical Electromagnetism

It is of interest to compare these results with a direct discretization of classical electromagnetism. Suppose that $X, X', X'', X''', \cdots$ is a time series of vectors in R^3 (where R denotes the real numbers). Let $dX = X' - X$ be the usual discrete derivative (with time step equal to one for convenience). Let $A \bullet B$ denote the usual inner product of vectors in three dimesions. Assume that there are fields E and B such that

$$d^2X = E + dX \times B$$

(the Lorentz force law). Assume also that E and B are perpendicular to the velocity vector dX, and that E is perpendicular to B. Then we have

$$dX' \times dX = (dX' - dX) \times dX = (d^2X) \times (dX)$$

$$= E \times dX + (dX \times B) \times dX$$
$$= E \times dX - dX(B \bullet dX) + (dX \bullet dX)B.$$

Since E is perpendicular to dX we know there is a λ such that $E \times dX = \lambda B$ and we have $B \bullet dX = 0$ since B is perpendicular to dX. Therefore

$$dX' \times dX = \lambda B + ||dX||^2 B$$

so that

$$B = dX' \times dX/(\lambda + ||dX||^2).$$

Up to the factor in the denominator, this formula is in exactly the same pattern as the formula in our discrete model for DOC electromagnetism as described in the previous subsection. To see this, note that the B field in the DOC model is proportional to $DX \times DX$ and that $DX = JdX$ so that $DX \times DX = JdX \times JdX = J^2 dX' \times dX$. Up to the time-shifting algebra and a proportionality constant, the expressions are the same! Clearly more work is needed in comparing classical discrete electromagnetism with the results of a discrete analysis of the Feynman-Dyson derivation.

3 Gauge Fields and Differential Geometry

Letting X_i $(i = 1, 2, ..., d)$ denote a set of spatial variables (non-commutative time series in the sense of our discrete ordered calculus), we will look at a collection of basic assumptions about the commutation of these variables and of their derivatives. It is natural from the point of view of the discrete ordered calculus to have

$$[X_i, X_j] = 0$$

for all i and j. There are no other natural commutations from the point of view of this calculus.

We shall define g_{ij} by the equation

$$[X_i, \dot{X}_j] = g_{ij}.$$

Here $\dot{X}_j$ is shorthand forDX_j and

$$[A, B] = AB - BA.$$

Along with this commutator equation, we will assume that

$$[X_i, X_j] = 0,$$

$$[X_i, g_{jk}] = 0$$

and

$$[g_{rs}, g_{jk}] = 0.$$

Here it is assumed that g_{ij} is non-degenerate in the sense that there exists g^{ij} so that

$$g^{ij} g_{jk} = \delta^i_k$$

and that

$$g_{ij} g^{jk} = \delta^k_i.$$

Here we are using the Einstein summation convention that implicitly assumes that we sum over repeated indices in an expression. Symbol δ^i_j is a Kronecker delta, equal to 1 when i equals j and 0 otherwise.

The first result that is a direct consequence of these assumptions is the symmetry of the "metric" coefficients g^{ij}. That is, we shall show that

$$g^{ij} = g^{ji}.$$

Lemma 3. $g_{ij} = g_{ji}$.

Proof.

$$\begin{aligned} & g_{ij} - g_{ji} \\ &= [X_i, \dot{X}_j] - [X_j, \dot{X}_i] \\ &= [X_i, \dot{X}_j] + [\dot{X}_i, X_j] \\ &= D[X_i, X_j] \\ &= 0. \end{aligned}$$

For the purpose of doing calculus in this situation we define $\dot{X}^i$ by the equation

$$\dot{X}^i = g^{ik}\dot{X}_k.$$

The operator $\dot{X}^i$ is simply the index shift of the corresponding $\dot{X}_i$. We do not define a corresponding X^i. It is easy to check the equation

$$[X_i, \dot{X}^j] = \delta_i^j.$$

Consequently, we define the derivative of an operator F with respect to X_i by the equation

$$\partial^i F = [F, \dot{X}^i]$$

and the corresponding lowered derivative by the formula

$$\partial_i F = [F, \dot{X}_i].$$

Note that we have

$$\partial_i X_j = g_{ij}.$$

We also define

$$\hat{\partial}_i F = [X_i, F],$$

the derivative of F with respect to the conjugate variable $\dot{X}^i$.

With these partial derivatives in hand, we define $\dot{F}$ by the formula

$$\dot{F} = \partial^k F \dot{X}_k.$$

If F commutes with g^{ij} then it is easy to see that

$$\dot{F} = \partial_k F \dot{X}^k.$$

These formulas extend (implicitly) the definition of the time series to entities other than the operators X_i since

$$\dot{F} = DF = J(F' - F).$$

A stream of consequences then follows by differentiating both sides of the equation

$$g_{ij} = [X_i, \dot{X}_j].$$

Note that

$$\dot{g_{ij}} = [\dot{X}_i, \dot{X}_j] + [X_i, D^2 X_j]$$

by the Leibniz rule

$$D[A, B] = [DA, B] + [A, DB].$$

Note also that we can freely use the Jacobi identity

$$[A, [B, C]] + [C, [A, B]] + [B, [C, A]] = 0.$$

In particular, the Levi-Civita connection

$$\Gamma_{ijk} = (1/2)(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})$$

associated with the g_{ij} comes up almost at once from the differentiation process described above. To see how this happens, view the following calculation where

$$\hat{\partial}_i \hat{\partial}_j F = [X_i, [X_j, F]].$$

We apply the operator $\hat{\partial}_i \hat{\partial}_j$ to the second DOC derivative of X_k.

Lemma 4. $\Gamma_{ijk} = (1/2)\hat{\partial}_i \hat{\partial}_j D^2 X_k$

Proof.

$$\hat{\partial}_i \hat{\partial}_j D^2 X_k = [X_i, [X_j, D^2 X_k]]$$

$$= [X_i, \dot{g_{jk}} - [\dot{X}_j, \dot{X}_k]]$$

$$= [X_i, \dot{g_{jk}}] - [X_i, [\dot{X}_j, \dot{X}_k]]$$

$$= [X_i, \dot{g_{jk}}] + [\dot{X}_k, [X_i, \dot{X}_j]] + [\dot{X}_j, [\dot{X}_k, X_i]$$

$$= [g_{jk}, \dot{X}_i] + [\dot{X}_k, g_{ij}] + [\dot{X}_j, -g_{ik}]$$

$$= \partial_i g_{jk} - \partial_k g_{ij} + \partial_j g_{ik}$$

$$= 2\Gamma_{kij}.$$

It is remarkable that the form of the Levi-Civita connection comes up directly from this non-commutative calculus without any apriori geometric interpretation. We shall discuss the context of this result in the next two sections of the paper.

One finds that

$$D^2 X_i = G_i + g_{ir} g_{js} F^{rs} \dot{X}^j + \Gamma_{ijk} \dot{X}^j \dot{X}^k$$

where

$$F^{rs} = [\dot{X}^r, \dot{X}^s].$$

It follows from the Jacobi identity that

$$F_{ij} = g_{ir} g_{js} F^{rs}$$

satisfies the equation

$$\partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij} = 0,$$

identifying F_{ij} as a non-commutative analog of a gauge field. G_i is a non-commutative analog of a scalar field. The details of these calculations will be found in [26].

This description of the equations for a non-commutative particle in a metric field illustrates the role of the background discrete time in this theory. In terms of the background time the metric coefficients are not constant. It is through this variation that the spacetime derivatives of the theory are articulated. The background is a process with its own form of discrete time, but no spacetime structure as we know and observe it. Our observation of spacetime structure appears as a rough (commutative) approximation to the processes described as consequences of the basic non-commutative equations of the discrete ordered calculus.

4 Curvature, Jacobi Identity and the Levi-Civita Connection

In this section, we go back to basics and examine the context of calculus defined via commutators. We shall use a partially index-free notation. In this notation, we avoid nested subscripts by using different variable names and then using these names as subscripts to refer to the relevant variables. Thus we write X and Y instead of X_i and X_j, and we write g_{XY} instead of g_{ij}. It is assumed that the derivation DX has the form $DX = [X, J]$ for some J.

The bracket $[A, B]$ is not assumed to be a commutator. It is assumed to satisfy the Jacobi identity, bilinearity in each variable, and the Leibniz rule for all functions of the form $\delta_K(A) = [A, K]$. That is we assume that

$$\delta_K(AB) = \delta_K(A)B + A\delta_K(B).$$

Recall that in classical differential geometry one has the notion of a covariant derivative, defined by taking a difference quotient using parallel translation via a connection. Covariant derivatives in different directions do not necessarily commute. The commutator of covariant derivatives gives rise to the curvature tensor in the form

$$[\nabla_i, \nabla_j]X^k = R^k_{lij}X^l.$$

If derivatives do not commute then we regard their commutator as expressing a curvature. In our non-commutative context this means that curvature arises *prior* to any notion of covariant derivatives since *even the basic derivatives do not commute.*

We shall consider derivatives in the form

$$\nabla_X(A) = [A, \Lambda_X].$$

Examine the following computation:

$$\nabla_X\nabla_Y F = [[F, \Lambda_Y], \Lambda_X] = -[[\Lambda_X, F], \Lambda_Y] - [[\Lambda_Y, \Lambda_X], F]$$

$$= [[F, \Lambda_X], \Lambda_Y] + [[\Lambda_X, \Lambda_Y], F]$$

$$= \nabla_Y \nabla_X F + [[\Lambda_X, \Lambda_Y], F].$$

Thus

$$[\nabla_X, \nabla_Y]F = R_{XY}F$$

where

$$R_{XY}F = [[\Lambda_X, \Lambda_Y], F].$$

We can regard R_{XY} as a curvature operator.

The analog in this context of flat space is abstract quantum mechanics! That is, we assume position variables (operators) $X, Y, \cdots$ and momentum variables (operators) $P_X, P_Y, \cdots$ satisfying the equations below.

$$[X, Y] = 0$$

$$[P_X, P_Y] = 0$$

$$[X, P_Y] = \delta_{XY}$$

where δ_{XY} is equal to one if X equals Y and is zero otherwise. We define

$$\partial_X F = [F, P_X]$$

and

$$\partial_{P_X} F = [X, F].$$

In the context of the above commutation relations, note that these derivatives behave correctly in that

$$\partial_X(Y) = \delta_{XY}$$

and

$$\partial_{P_X}(P_Y) = \delta_{XY}$$

$$\partial_{P_X}(Y) = 0 = \partial_X(P_Y)$$

with the last equations valid even if $X = Y$. Note also that iterated partial derivatives such as $\partial_X \partial_Y$ commute. Hence the curvature R_{XY} is equal to zero. We shall regard these position and momentum operators and the corresponding partial derivatives as an abstract algebraic substitute for flat space.

With this reference point of (algebraic, quantum) flat space we can define

$$\hat{P}_X = P_X - A_X$$

for an arbitrary algebra-valued function of the variable X. In indices this would read

$$\hat{P}_i = P_i - A_i,$$

and with respect to this deformed momentum we have the covariant derivative

$$\nabla_X F = [F, \hat{P}_Y] = [F, P_Y + A_Y] = \partial_Y F + [F, A_Y].$$

The curvature for this covariant derivative is given by the formula

$$R_{XY}F = [\nabla_X, \nabla_Y]F = [[\lambda_X, \lambda_Y], F]$$

where $\lambda_X = P_X - A_X$. Hence

$$R_{XY} = [P_X - A_X, P_Y - A_Y] = -[P_X, A_Y] - [A_X, P_Y] + [A_X, A_Y]$$

$$= \partial_X A_Y - \partial_Y A_X + [A_X, A_Y].$$

With indices this reads

$$R_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j].$$

and the reader will note that this has the abstract form of the curvature of a Yang-Mills gauge field, and specifically the form of the electromagnetic field when the potentials A_i and A_j commute with one another.

Continuing with this example, we compute

$$[X, \hat{P}_Y] = [X, P_Y - A_Y] = \delta_{XY} - [X, A_Y].$$

Let

$$g_{XY} = \delta_{XY} - [X, A_Y]$$

so that

$$[X, \hat{P}_Y] = g_{XY}.$$

We will shortly consider the form of this general case, but first it is useful to restrict to the case where $[X, A_Y] = 0$ so that $g_{XY} = \delta_{XY}$. This is the domain

to which the original Feynman-Dyson derivation applies. In order to enter this domain, we set

$$\dot{X} = DX = \hat{P}_X = P_X - A_X.$$

We then have

$$[X_i, X_j] = 0$$
$$[X_i, \dot{X}_j] = \delta_{ij}$$

and

$$R_{ij} = [\dot{X}_i, \dot{X}_j] = \partial_i A_j - \partial_j A_i + [A_i, A_j].$$

Note that even under these restrictions we are still looking at the possibility of a non-abelian gauge field. The pure electromagnetic case is when the commutator of A_i and A_j vanishes. But why do we set $\dot{X} = \hat{P}_X$? The answer to this is the key to the gauge interpretation of electromagnetism, for with this interpretation we find that $\dot{X}$ satisfies the Lorentz force law $\ddot{X} = E + \dot{X} \times B$ where B represents the magnetic field and E the electric field (in the case of three space variables X_i with $i = 1, 2, 3$.) To see how this works, suppose that $\ddot{X}_i = E_i + F_{ij}\dot{X}_j$ and suppose that E_i and F_{ij} commute with X_k. Then we can compute

$$[X_i, \ddot{X}_j] = [X_i, E_j + F_{jk}\dot{X}_k]$$
$$= F_{jk}[X_i, \dot{X}_k] = F_{jk}\delta_{ik} = F_{ji}.$$

This implies that

$$F_{ij} = [\dot{X}_i, \dot{X}_j] = R_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]$$

since $[X_i, \ddot{X}_j] + [\dot{X}_i, \dot{X}_j] = D[X_i, \dot{X}_j] = 0$. It is then easy to verify that the Lorentz force equation is satisfied with $B_k = \epsilon_{ijk}R_{ij}$ and that in the case of $[A_i, A_j] = 0$ this leads directly to standard electromagnetic theory when the bracket is a Poisson bracket (see the next section for a discussion of Poisson brackets). When this bracket is not zero but the potentials A_i are functions only of the X_j we can look at a generalization of gauge theory where the non-commutativity comes from internal Lie algebra parameters. This shows that the Feynman-Dyson derivation supports certain generalizations of classical electromagnetism, and this will be the subject of a more expanded version of this paper.

In regard to this last remark, the reader should note that in our [28, 27] algebraic and discrete version of the Feynman-Dyson derivation it was actually an additional assumption that $B \times B = 0$ where $B \times B$ denotes the (non-commutative) vector cross product of B with itself. (Note that $B = (1/2)\dot{X} \times \dot{X}$.) In the original Dyson paper this cross product vanished because of assumptions about the operators and their Hilbert space representations. With $B \times B$ as an extra term, the Feynman-Dyson derivation is indeed a non-commutative generalization of electromagnetism and includes forms of gauge theories among its models.

Generalizing, we wish to examine the structure of the following special axioms for a bracket.

$$[X, DY] = g_{XY}$$

$$[X, Y] = 0$$

$$[Z, g_{XY}] = 0$$

$$[g_{XY}, g_{ZW}] = 0$$

Note that

$$Dg_{YZ} = D[Y, DZ] = [DY, DZ] + [Y, D^2 Z].$$

and that $D[X, g_{XY}] = 0$ implies that

$$[g_{XY}, DZ] = [Z, Dg_{XY}].$$

Define two types of derivations as follows

$$\nabla_X(F) = [F, DX]$$

and

$$\nabla_{DX}(F) = [X, F].$$

These are dual with respect to g_{XY} and will act like partials with respect to these variables in the special case when g_{XY} is a Kronecker delta, δ_{XY}. If the form g_{XY} is invertible, then we can rewrite these derivations by contracting the inverse of g to obtain standard formal partials.

$$\nabla_{DX}\nabla_{DY}D^2Z = [X,[Y,D^2Z]]$$
$$= [X, Dg_{YZ} - [DY,DZ]] = [X, Dg_{YZ}] - [X,[DY,DZ]]$$
$$= [g_{YZ}, DX] - [X,[DY,DZ]]$$
$$= \nabla_X(g_{YZ}) - [X,[DY,DZ]].$$

Now use the Jacobi identity on the second term and obtain

$$\nabla_{DX}\nabla_{DY}D^2Z = \nabla_X(g_{YZ}) + [DZ,[X,DY]] + [DY,[DZ,X]]$$
$$= \nabla_X(g_{YZ}) - \nabla_Z(g_{XY}) + \nabla_Y(g_{XZ}).$$

This is the formal Levi-Civita connection.

At this stage we face once again the mystery of the appearance of the Levi-Civita connection. There is a way to see that the appearance of this connection is not an accident, but rather quite natural. We shall explain this point of view in the next section where we discuss Poisson brackets and the connection of this formalism with classical physics. On the other hand, we have seen in this section that it is quite natural for curvature in the form of the non-commutativity of derivations to appear at the outset in a non-commutative formalism. We have also see that this curvature and connection can be understood as a measurement of the deviation of the theory from the "flat" commutation relations of ordinary quantum mechanics. Electromagnetism and Yang-Mills theory can be seen as the theory of the curvature introduced by such a deviation. On the other hand, from the point of view of metric differential geometry, the Levi-Civita connection is the unique connection that preserves the inner product defined by the metric under the parallel translation defined by the connection. We would like to see that the formal Levi-Civita connection produced here has this property as well.

To this end lets recall the formalism of parallel translation. The infinitesimal parallel translate of A is denoted by $A' = A + \delta A$ where

$$\delta A^k = -\Gamma^k_{ij} A^i dX^j$$

where here we are writing in the usual language of vectors and differentials with the Einstein summation convention for repeated indices. We assume

that the Christoffel symbols satisfy the symmetry condition $\Gamma^k_{ij} = \Gamma^k_{ji}$. The inner product is given by the formula

$$< A, B >= g_{ij}A^iB^j$$

Note that here the bare symbols denote vectors whose coordinates may be indicated by indices. The requirement that this inner product be invariant under parallel displacement is the requirement that $\delta(g_{ij}A^iA^j) = 0$. Calculating, one finds

$$\delta(g_{ij}A^iA^j) = (\partial_k g_{ij})A^iA^j dX^k + g_{ij}\delta(A^i)A^j + g_{ij}A^i\delta(A^j)$$

$$= (\partial_k g_{ij})A^iA^j dX^k - g_{ij}\Gamma^i_{rs}A^r dX^s A^j - g_{ij}A^i\Gamma^j_{rs}A^r dX^s$$

$$= (\partial_k g_{ij})A^iA^j dX^k - g_{ij}\Gamma^i_{rs}A^r A^j dX^s - g_{ij}\Gamma^j_{rs}A^i A^r dX^s$$

$$= (\partial_k g_{ij})A^iA^j dX^k - g_{sj}\Gamma^s_{ik}A^iA^j dX^k - g_{is}\Gamma^s_{jk}A^iA^j dX^k$$

Hence

$$(\partial_k g_{ij}) = g_{sj}\Gamma^s_{ik} + g_{is}\Gamma^s_{jk}.$$

From this it follows that

$$\Gamma_{ijk} = g_{is}\Gamma^s_{jk} = (1/2)(\partial_k g_{ij} - \partial_i g_{jk} + \partial_j(g_{ik})).$$

Certainly these notions of variation can be imported into our abstract context. The question remains how to interpret the new connection that arises. We now have a new covariant derivative in the form

$$\hat{\nabla}_i X^j = \partial_i X^j + \Gamma^j_{ki}X^k.$$

The question is how the curvature of this connection interfaces with the gauge potentials that gave rise to the metric in the first place. The theme of this investigation has the flavor of gravity theories with a gauge theoretic background. We will investigate these relationships in detail in a sequel to this paper.

5 Poisson Brackets and Commutator Brackets

Dirac [11] introduced a fundamental relationship between quantum mechanics and classical mechanics that is summarized by the maxim *replace Poisson brackets by commutator brackets.* Recall that the Poisson bracket $\{A, B\}$ is defined by the formula

$$\{A, B\} = (\partial A/\partial q)(\partial B/\partial p) - (\partial A/\partial p)(\partial B/\partial q),$$

where q and p denote classical position and momentum variables respectively.

In our version of discrete physics the noncommuting variables are functions of discrete time, with a DOC derivative D as described in the first section. Since $DX = XJ - JX = [X, J]$ is itself a commutator, it follows that

$$D([A, B]) = [DA, B] + [A, DB]$$

for any expressions A, B in our ring R. A corresponding Leibniz rule for Poisson brackets would read

$$(d/dt)\{A, B\} = \{dA/dt, B\} + \{A, dB/dt\}.$$

However, here there is an easily verified exact formula:

$$(d/dt)\{A, B\} = \{dA/dt, B\} + \{A, dB/dt\} - \{A, B\}(\partial\dot{q}/\partial q + \partial\dot{p}/\partial p).$$

This means that the Leibniz formula will hold for the Poisson bracket exactly when

$$(\partial\dot{q}/\partial q + \partial\dot{p}/\partial p) = 0.$$

This is an integrability condition that will be satisfied if p and q satisfy Hamilton's equations

$$\dot{q} = \partial H/\partial p,$$

$$\dot{p} = -\partial H/\partial q.$$

This, of course, means that q and p are following a principle of least action with respect to the Hamiltonian H. Thus we can interpret the *fact* $D([A,B]) = [DA,B] + [A,DB]$ in the discrete (commutator) context as an analog of the principle of least action. Taking the discrete context as fundamental, we say that Hamilton's equations are *motivated* by the presence of the Leibniz rule for the discrete derivative of a commutator. The classical laws are obtained by following Dirac's maxim in the opposite direction! Classical physics is produced by following the correspondence principle upwards from the discrete.

Taking the last paragraph seriously, we must reevaluate the meaning of Dirac's maxim. The meaning of quantization has long been a basic mystery of quantum mechanics. By traversing this territory in reverse, starting from the non-commutative world, we begin these questions anew.

In making this backwards journey to classical physics we see how our earlier assertion that bare quantum mechanics of commutators can be regarded as the background for the coupling with other fields (as in the description of formal gauge theory in the last section), fits with Poisson brackets. The bare Poisson brackets satisfy

$$\{q_i, q_j\} = 0$$

$$\{p_i, p_j\} = 0$$

$$\{q_i, p_j\} = \delta_{ij}.$$

In our previous formalism, we would identify X_i as the correspondent with q_i and P_j as the correspondent of p_j. And, given a classical vector potential A, we could write the coupling $dq_i/dt = p_i - A_i$ to describe the motion of a particle in the presence of an electromagnetic field. The analog of the Feynman Dyson derivation is then expressed classically in terms of the Poisson brackets. Similar remarks apply to the analogs for gauge theory and curvature. In particular it is of interest to see that our derivation of the Levi-Civita connection corresponds to the motion of a particle in generalized coordinates that satisfies Hamilton's equations. The fact that such a particle moves in a geodesic according to the Levi-Civita connection is a classical fact that was surely one of the motivations for the development of differential geometry. Our derivation of the Levi-Civita connection, interpreted in Poisson brackets, reproduces this result.

To see how this works, let $ds^2 = g^{ij}dx_i dx_j$ denote the metric in the generalized coordinates x_k. Then the velocity of the particle has square $v^2 = (ds/dt)^2 = g^{ij}\dot{x}_i\dot{x}_j$. The Lagrangian for the system is the kinetic energy $L = mv^2/2 = mg^{ij}\dot{x}_i\dot{x}_j/2$. Then the canonical momentum is $p_j = \partial L/\partial\dot{x}_j$, and with $q_i = x_i$ we have the Poisson brakets

$$\delta_{ij} = \{q_i, p_j\} = \{x_i, \partial L/\partial\dot{x}_j\} = \{x_i, mg^{jk}\dot{x}_k\}.$$

Taking $m = 1$ for simplicity, we can rewrite this bracket as

$$\{x_i, \dot{x}_j\} = g_{ij}.$$

This, in Poisson brackets, is our generalized equation of motion.

The classical derivation applies Lagrange's equation of motion to the system. Lagrange's equation reads

$$d/dt(\partial L/\partial\dot{x}_i) = \partial L/\partial x_i.$$

Since this equation is equivalent to Hamilton's equation of motion, it follows that the Poisson brackets satisfy the Leibniz rule. With this, we can proceed with our derivation of the Levi-Civita connection in relation to the acceleration of the particle. In the classical derivation, one writes out the Lagrange equation and solves for the acceleration. The advantage of using only the Poisson brackets is that it shows the relationship of the connection with the Jacobi identity and the Leibniz rule.

This discussion raises further questions about the nature of the generalization that we have made. Originally Hermann Weyl [44] generalized classical differential geometry and discovered gauge theory by allowing changes of length as well as changes of angle to appear in the holonomy. Here we arrive at a very similar situation via the properties of a non-commutative discrete calculus of observations. A closer comparison with the geometry of gauge theories is called for.

6 Discussion on q-Deformation

The direct relation between the content of local physical descriptions based on the *DOC* calculus and more global considerations are a matter of speculation.

One strong hint is contained in the properties of the discrete derivative that has the form

$$D_q f(x) = (f(qx) - f(x))/(qx - x).$$

The classical derivative occurs in the limit as q approaches one.

In the setting of q not equal to one, the derivative D_q is directly related to fundamental noncommutativity. Consider variables x and y such that $yx = qxy$ where q is a commuting scalar. Then the expansion of $(x+y)^n$ generates a q-binomial theorem with q-choice coefficients composed in q-factorials of q-integers $[n]_q$ where

$$[n]_q = 1 + q + q^2 + ... + q^{(n-1)}.$$

The derivative D_q is directly related to the q-integers via the formula

$$D_q(x^n) = [n]_q x^{n-1}.$$

In the context of this paper, we have considered discrete derivatives in the form

$$d_\Delta f(x) = (f(x+\Delta) - f(x))/\Delta.$$

This will convert to the q-derivative if $x + \Delta = qx$. Thus we need

$$q = (x + \Delta)/x.$$

This means that a direct translation from DOC to q-derivations could be effected if we allowed q to vary as a function of x and introduced the temporal operator J into the calculus of q-derivatives.

In general, many q-deformed structures such as the quantum groups associated with the classical Lie algebras appear to be entwined with the discretization inherent in D_q. The quantum groups have turned out to be deeply connected with topological amplitudes for networks describing knots and three dimensional spaces. (See the next section of this paper.) The analog for the quantum groups in dimension four is being sought. If there is a connection between the local and the global parts of our essay it may lie in hidden connections between discretization and quantum groups. Clearly there is much work to be done in this field.

There is a clue about the meaning of the operator J ($DF = [F, J]$ in the discrete ordered calculus) in the context of quantum groups. Quantum groups are Hopf algebras. A quantum group such as $G = U_q(SU(2))$ is actually an algebra over a field k with an antipode

$$S : G \longrightarrow G$$

and a coproduct

$$\Delta : G \longrightarrow G \otimes G,$$

a unit 1 and a couinit

$$\epsilon : G \longrightarrow k.$$

The coproduct is a map of algebras. The antipode is an antimorphism, $S(xy) = S(y)S(x)$, and generalizes the inverse in a group in the sense that $\Sigma S(x_1)x_2 = \epsilon(x)1$ and $\Sigma x_1 S(x_2) = \epsilon(x)1$ where $\Delta(x) = \Sigma x_1 \otimes x_2$.

An element g in a quantum group G is said to be a *grouplike element* if $\Delta(g) = g \otimes g$ and $S(g) = g^{-1}$. In many quantum groups (such as $G = U_q(SU(2))$) the square of the antipode is represented via conjugation by a special grouplike element that we shall denote by J. Thus

$$S^2(x) = J^{-1}xJ$$

for all x in G. This means that it is possible to define the discrete ordered calculus in the context of a quantum group G (as above) by taking J to be the special grouplike element. Then we have

$$DX = [X, J] = XJ - JX = J(J^{-1}XJ - X) = J(S^2(X) - X).$$

Conjugation by the special grouplike element in the quantum group constitutes the time evolution operator in this algebra.

There are a number of curious aspects to this use of the discrete ordered calculus in a quantum group. First of all, it is the case that in some quantum groups (for example with undeformed classical Lie algebras) the square of the antipode is equal to the identity mapping. From the point of view of *DOC*, time does not exist in these algebras. But in the q-deformations such as $U_q(SU(2))$, the square of the antipode is quite non-trivial and can serve well as the tick of the clock. In this way, q-deformations do provide a context for time. In particular, this suggests that the q-deformations of classical spin

networks [38] should be able to accommodate time. A suggestion directly related to this remark occurs in [9], and we shall take this up at the end of the next section of this paper.

7 Networks, Discrete Spacetime and the Dirac Equation

One can consider replacing continuous space (such as Euclidean space with the usual topology) by a discrete structure of relationships. The geometry of the Greeks held a discrete web of relationships in the context of continuous space. That space was not coordinatized in our way, nor was it held as an infinite aggregate of points. In general topology there is a wide choice for possible spatial structures (where we mean by a space a topology on some set).

Discretization of space and time implicates the replacement of spacetime by a network, graph or complex that has nodes for the points and edges to indicate significant relationships among the points.

Euler's work in the eighteenth century brought forth the use of abstract graphs as holders of spatial structure. After Euler it was possible to find the classification of the Greek regular solids in the the (wider) classification of the regular graphs on the surface of the sphere. Metric can disappear into relationship under the topological constraint of Euler's formula $V-E+F=2$, where V denotes the number of vertices, E the number of edges and F the number of faces for the connected graph G on the sphere.

A network itself can represent an abstract space. Embeddings of that network into a given space (such as graphs on the two dimensional sphere) correspond to global constraints on the structure of the abstract graph.

Now a new theme arises, motivated by a conjunction of combinatorics and physics. Imagine labelling the edges of the network from some set of "colors". These colors can represent the basic states of a physical system, or they can be an abstract set of distinct markers for purely mathematical purposes. Once the network is labelled, each vertex is an entity with a collection of labels incident to it. Let there be given a function that associates a number (or algebra element) to each such labelled vertex. Call this number the *vertex*

weight at that vertex. Let C denote a specific coloring of the network N and consider the product, over all the vertices of N of the values of the vertex weights. Finally let $Z(N)$, the *amplitude* of the network, be defined as the summation of the product of the vertex weights over all colorings of the net. $Z(N)$ is also called the *partition function* of the network.

Amplitudes of this sort are exactly what one computes in finding the partition function of a physical system or the quantum mechanical amplitude for a discrete process. In all these cases the network is interwoven with the algebraic structure of the vertex weights. It is only recently that topological properties of networks in three dimensional space have come to be understood in this way [22], [1],[45]. This has led to new information about the topology of low dimensional spaces, and new relationships between physics and topology.

A classical example of such an amplitude was discovered by Roger Penrose [5] in elucidating special colorings of 3-regular graphs in the plane. A 3-regular graph G has three edges incident to each vertex. When embedded in the plane, these edges acquire a specific cyclic order. Three colors are used. One associates to each vertex the weight

$$\sqrt{-1}\,\epsilon_{abc}$$

where a,b,c denote the edges meeting the vertex in this cyclic order, and the epsilon is equal to 1, -1 according as the edges have distinct labels in the given or reverse cyclic order, or 0 if there is a repetition of labels. The resulting amplitude counts the number of ways to color the network with three colors so that three distinct colors are incident to each vertex. This result is a perspicuous generalization of the classical four color problem of coloring maps in the plane with four colors so that adjacent regions receive different colors.

The Penrose example generalizes to networks whose amplitudes embody geometrical properties of Euclidean three dimensional space (angles and their dependence). Geometry begins to emerge in terms of the averages of properties of an abstract and discrete network of relationships. Topological properties emerge in the same way. The idea of space may change to the idea of a network with global states and a functor that associates this network and its states to the more familiar properties that a classical observer might see.

7.1 Remarks on Quantum Mechanics

We should remark on the basic formalism for amplitudes in quantum mechanics. The Dirac notation $\langle A|B\rangle$ [11] denotes the probability amplitude for a transition from A to B. Here A and B could be points in space (for the path of a particle), fields (for quantum field theory), or geometries on spacetime (for quantum gravity). The probability amplitude is a complex number. The actual probability of an event is the absolute square of the amplitude. If a complete set of intermediate states $C_1, C_2, ...C_n$ is known, then the amplitude can be expanded to a summation

$$\langle A|B\rangle = \Sigma_{i=1}^{n} \langle A|C_i\rangle\langle C_i|B\rangle.$$

This formula follows the formalism of the usual rules for probability, and it allows for the constructive and destructive interference of the amplitudes. It is the simplest case of a quantum network of the form

$$A - - - * - - - C - - - * - - - B$$

where the colors at A and B are fixed and we run through all choices of colors for for the middle edge. The vertex weights at the vertices labelled $*$ are $\langle A|C\rangle$ and $\langle C|B\rangle$ respectively. A measurement at the C edge reduces the big summation to a single value.

Consider the generalization of the previous example to the graph

$$A - - - * - - - C^1 - - - * - - - C^2 - - - * - - - ... - - - * - - - C^m - - - B$$

With A and B fixed the amplitude for the net is

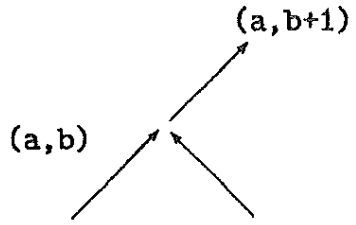
$$< A|B >= \Sigma_{1 \leq i_1 \leq ... \leq i_m \leq n} < A|C^1_{i_1} >< C^2_{i_2}|C^3_{i_3} > ... < C^m_{i_m}|B >$$

One can think of this as the sum over all the possible paths from A to B. In fact in the case of a "particle" travelling between two points in space, this is exactly what must be done to compute an amplitude - integrate over all the paths between the two points with appropriate weightings. In the discrete case this sort of summation makes perfect sense. In the case of a continuum there is no known way to make rigorous mathematical sense out of all cases of such integrals. Nevertheless, the principles of quantum

mechanics must be held foremost for physical purposes and so such "path integrals" and their generalizations to quantum fields are in constant use by theoretical physicists [16] who take the point of view that the proof of a technique is in the consistency of the results with the experiments. When the observations themselves are mathematical (such as finding invariants of knots and links), the issue acquires a new texture.

Now consider the summation discussed above in the case where $n = 2$. That is, we shall assume that each C^kcan take two values, call these values L and R. Furthermore let us suppose that $< L|R >=< R|L >= \sqrt{-1}$ while $< L|L >=< R|R >= 1$. The amplitudes that one computes in this case correspond to solutions to the Dirac equation [11] in one space variable and one time variable. This example is related to an observation of Richard Feynman [16]. In [25] we give a very elementary derivation of this result and we show how these amplitudes give solutions to the discretized Dirac equation, so everything is really quite exact and one can understand just what happens in taking the limit to the continuum. In this example a state of the network consists in a sequence of choices of L or R. These can be interpreted as choices to move left or right along the light-cone in a Minkowski plane. It is in summing over such paths in spacetime that the solution to the Dirac equation appears. In this case, time has been introduced into the net by interpreting the sequence of nodes in the network as a temporal direction.

More specifically, let (a, b) denote a point in discrete Minkowski spacetime in lightcone coordinates. This means that a denotes the number of steps taken to the left and b denotes the number of steps taken to the right. We let $\psi_L(a, b)$ denote the sum over the paths that enter the point (a, b) from the left and $\psi_R(a, b)$ the sum over the paths that enter (a, b) from the right. Each path P contributes $i^{c(P)}$where $c(P)$ denotes the number of corners in the path. View the diagram below.



It is clear from the diagram that

$$\psi_L(a, b+1) = \psi_L(a, b) + i\psi_R(a, b).$$

Thus we have that

$$\partial\psi_L/\partial R = i\psi_R$$

and similarly

$$\partial\psi_R/\partial L = i\psi_L.$$

This pair of equations is the Dirac equation in light cone coordinates.

This discrete derivation of the Dirac equation is simpler than the method used in [25]. I am indebted to Charles Bloom [3] for pointing this out to me. In fact, this form of the discretization is essentially Feynman's original method as is evident from the reproduction of Feynman's handwritten notes in Figure 8 of the review paper [39] by Schweber. For one approach, very close in spirit, that generalizes this exercise of Feynman to four dimensional discrete spacetime see [40].

As in the Dirac equation example, one way to incorporate spacetime is to introduce a temporal direction into the net. At a vertex, one must specify labels of *before* and *after* to each edge of the net that is incident to that vertex. If there is a sufficiently coherent assignment of such local times, then a global time direction can emerge for the entire network. Networks endowed with temporal directions have the structure of morphisms in a category where each morphism points from past to future. A category of quantum networks emerges equipped with a functor (via the algebra of the vertex weights) to morphisms of vector spaces and representations of generalized symmetry groups. Appropriate traces of these morphisms produce the amplitudes.

Quantum non-locality is built into the network picture. Any observer taking a measurement in the net has an effect on the global set of states available for summation and hence affects the possibilities of observations at all other nodes in the network. By replacing space with a network we obtain a precursor to spacetime in which quantum mechanics is built into the initial structure.

Remark. A striking parallel to the views expressed in this section can be found in [12]. Concepts of time and category are discussed by Louis Crane [8], [9] in relation to topological quantum field theory. In the case of Crane's

work there is a deeper connection with the methods of this paper, as I shall explain below.

7.2 Temporality and the Crane Model for Quantum Gravity

Crane uses a partition function defined for a triangulated four-manifold. Let us denote the partition function by $Z(M^4, A, B) =< A|B >_M$ where M^4 is a four-manifold and A and B are (colored - see the next sentence) three dimensional submanifolds in the boundary of M. The partition function is constructed by summing over all colorings of the edges of a dual complex to this triangulation from a finite set of colors that correspond to certain representations of the the quantum group $U_q(SU(2))$ where q is a root of unity. The sum is over products of $15J_q$ symbols (natural generalizations of the $6J$ symbols in angular momentum theory) evaluated with respect to the colorings. The specific form of the partition function (here written in the case where A and B are empty) is

$$Z(M^4) = N^{v-e}\Sigma_\lambda \Pi_\sigma dim_q(\lambda(\sigma))\Pi_\tau dim_q^{-1}(\lambda(\tau))\Pi_\zeta 15J_q(\lambda(\zeta)).$$

Here λ denotes the labelling function, assigning colors to the faces and tetrahedra of M^4 and $v - e$ is the difference of the number of vertices and the number of edges in M^4. Faces are denoted by σ, tetrahedra by τ and 4-simplices by ζ. We refer the reader to [6] for further details.

In computing $Z(M^4, A, B) =< A|B >_M$ one fixes the choice of coloration on the boundary parts A and B. The analog with quantum gravity is that a colored three manifold A can be regarded as a three manifold with a choice of (combinatorial) metric. The coloring is the combinatorial substitute for the metric. In the three manifold case this is quite specifically so, since the colors can be regarded as affixed to the edges of the simplices. The color on a given edge is interpreted as the generalized distance between the endpoints of the edge. Thus $< A|B >_M$ is a summation over "all possible metrics" on M^4 that can extend the given metrics on A and B. $< A|B >_M$ is an amplitude for the metric (coloring) on A to evolve in the spacetime M^4 to the metric (coloring) on B.

The partition function $Z(M^4, A, B) =< A|B >_M$ is a topological invariant of the four manifold M^4. In particular, if A and B are empty (a vacuum-vacuum amplitude), then the Crane-Yetter invariant, $Z(M^4)$, is a function of the signature and Euler characteristic of the four-manifold [6]. On the mathematical side of the picture this is already significant since it provides a new way to express the signature of a four-manifold in terms of local combinatorial data.

From the point of view of a theory of quantum gravity, $Z(M^4, A, B) =< A|B >_M$, as we have described it so far, is lacking in a notion of time and dynamical evolution on the four manifold M^4. One can think of A and B as manifolds at the initial and final times, but we have not yet described a notion of time within M^4 itself.

Crane proposes to introduce time into M^4 and into the partition function $< A|B >_M$ by labelling certain three dimensional submanifolds of M^4 with special grouplike elements from the quantum group $U_q(SU(2))$ and extending the partition function to include this labelling. Movement across such a labelled hypersurface is regarded as one tick of the clock. The special grouplike elements act on the representations in such a way that the partition function can be extended to include the extra labels. Then one has the project to understand the new partition function and its relationship with discrete dynamics for this model of quantum gravity.

Lets denote the special grouplike element in the Hopf algebra $G = U_q(SU(2))$ by the symbol J. Then, as discussed at the end of the previous section, one has that the square of the antipode $S : G \longrightarrow G$ is given by the formula $S^2(x) = J^{-1}xJ$. This is the tick of the clock. The DOC derivative in the quantum group is given by the formula $DX = [X, J] = J(S^2(X) - X)$. I propose to generalize the discrete ordered calculus on the quantum group to a discrete ordered calculus on the four manifold M^4 with its hyperthreespaces labelled with special grouplikes. This generalised calculus will be a useful tool in elucidating the dynamics of Crane's model. Much more work needs to be done in this domain.

8 Appendix on Iterants

The primitive idea behind an iterant is a periodic time series or "waveform"

$$\cdots ababababab \cdots .$$

The elements of the waveform can be any mathematically or empirically well-defined objects. We can regard the ordered pairs $[a, b]$ and $[b, a]$ as abbreviations for the waveform or as two points of view about the waveform (a first or b first). Call $[a, b]$ an *iterant*. One has the collection of transformations of the form $T[a, b] = [ka, k^{-1}b]$ leaving the product ab invariant. This tiny model contains the seeds of special relativity, and the iterants contain the seeds of general matrix algebra! Since this paper has been a combination of discussions of non-commutativity and time series, we include this appendix on iterants. A more complete discussion will appear elsewhere. For related discussion see [18, 19, 20, 21, 23, 31, 32, 41].

Define products and sums of iterants as follows

$$[a, b][c, d] = [ac, bd]$$

and

$$[a, b] + [c, d] = [a + c, b + d].$$

The operation of juxtapostion is multiplication while + denotes ordinary addition in a category appropriate to these entities. These operations are natural with respect to the structural juxtaposition of iterants:

$$...ababababab...$$

$$...cdcdcdcdcdcd...$$

Structures combine at the points where they correspond. Waveforms combine at the times where they correspond. Iterants combine in juxtaposition.

If $\bullet$ denotes any form of binary compositon for the ingredients (a,b,...) of iterants, then we can extend $\bullet$ to the iterants themselves by the definition $[a, b] \bullet [c, d] = [a \bullet c, b \bullet d]$. In this section we shall first apply this idea to Lorentz transformations, and then generalize it to other contexts.

So, to work: We have

$$[t - x, t + x] = [t, t] + [-x, x] = t[1, 1] + x[-1, 1].$$

Since $[1,1][a,b] = [1a, 1b] = [a,b]$ and $[0,0][a,b] = [0,0]$, we shall write

$$1 = [1,1]$$

and

$$0 = [0,0].$$

Let

$$\sigma = [-1,1].$$

σ is a significant iterant that we shall refer to as a *polarity*. Note that

$$\sigma\sigma = 1.$$

Note also that

$$[t-x, t+x] = t + x\sigma.$$

Thus the points of spacetime form an algebra analogous to the complex numbers whose elements are of the form $t + x\sigma$ with $\sigma\sigma = 1$ so that

$$(t + x\sigma)(t' + x'\sigma) = tt' + xx' + (tx' + xt')\sigma.$$

In the case of the Lorentz transformation it is easy to see the elements of the form $[k, k^{-1}]$ translate into elements of the form

$$T(v) = [(1+v)/\sqrt{(1-v^2)}, (1-v)/\sqrt{(1-v^2)}] = [k, k^{-1}].$$

Further analysis shows that v is the relative velocity of the two reference frames in the physical context. Multiplication now yields the usual form of the Lorentz transform

$$T_k(t + x\sigma) = T(v)(t + x\sigma)$$

$$= (1/\sqrt{(1-v^2)} - v\sigma/\sqrt{(1-v^2)})(t + x\sigma)$$

$$= (t - xv)/\sqrt{(1-v^2)} + (x - vt)\sigma/\sqrt{(1-v^2)}$$

$$= t' + x'\sigma.$$

The algebra that underlies this iterant presentation of special relativity is a relative of the complex numbers with a special element σ of square one rather than minus one ($i^2 = -1$ in the complex numbers).

The appearance of a square root of minus one unfolds naturally from iterant considerations. Define the "shift" operator D on iterants by the equation

$$D[a,b] = [b,a].$$

Sometimes it is convenient to think of D as a delay opeator, since it shifts the waveform *...ababab...* by one internal time step. Now define

$$i[a,b] = \sigma D[a,b] = [-1,1][b,a] = [-b,a].$$

We see at once that

$$ii[a,b] = [-a,-b] = [-1,-1][a,b] = (-1)[a,b].$$

Thus

$$ii = -1.$$

This is the traditional construction of the square root of minus one in terms of operations on ordered pairs. Here we have described $i[a,b]$ in a *new* way as the superposition of the waveforms $\sigma = [-1,1]$ and $D[a,b]$ where $D[a,b]$ is the delay shift of the waveform $[a,b]$.

8.1 MATRIX ALGEBRA VIA ITERANTS

Matrix algebra has some strange wisdom built into its very bones. Consider a two dimensional periodic pattern or "waveform."

.....

...abababababababab...

...cdcdcdcdcdcdcdcd...

...abababababababab...

...cdcdcdcdcdcdcdcd...

...abababababababab...

.....

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} b & a \\ d & c \end{pmatrix}, \begin{pmatrix} c & d \\ a & b \end{pmatrix}, \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

Above are some of the matrices apparent in this array. Compare the matrix with the "two dimensional waveform" shown above. A given matrix freezes out a way to view the infinite waveform. In order to keep track of this patterning, lets write

$$[a,d] + [b,c]\eta = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

where

$$[x,y] = \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}.$$

and

$$\eta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The four matrices that can be framed in the two-dimensional wave form are all obtained from the two iterants $[a,d]$ and $[b,c]$ via the delay shift operation $D[x,y] = [y,x]$ which we shall denote by an overbar as shown below

$$D[x,y] = \overline{[x,y]} = [y,x].$$

Letting $A = [a,d]$ and $B = [b,c]$, we see that the four matrices seen in the grid are

$$A + B\eta, B + A\eta, \overline{B} + \overline{A}\eta, \overline{A} + \overline{B}\eta.$$

The operator η has the effect of rotating an iterant by ninety degrees in the formal plane. Ordinary matrix multiplication can be written in a concise form using the following rules:

$$\eta\eta = 1$$
$$\eta Q = \overline{Q}\eta$$

where Q is any two element iterant.
For example, let $\epsilon = [-1,1]$ so that $\overline{\epsilon} = -\epsilon$ and $\epsilon\epsilon = [1,1] = 1$. Let

$$i = \epsilon\eta.$$

Then

$$ii = \epsilon\eta\epsilon\eta = \epsilon\overline{\epsilon}\eta\eta = \epsilon(-\epsilon) = -\epsilon\epsilon = -1.$$

We have reconstructed the square root of minus one in the form of the matrix

$$i = \epsilon\eta = [-1, 1]\eta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

More generally, we see that

$$(A + B\eta)(C + D\eta) = (AC + B\overline{D}) + (AD + B\overline{C})\eta$$

writing the 2×2 matrix algebra as a system of hypercomplex numbers. Note that

$$(A + B\eta)(\overline{A} - B\eta) = A\overline{A} - B\overline{B}$$

The formula on the right corresponds to the determinant of the matrix. Thus we define the *conjugate* of $A + B\eta$ by the formula

$$\overline{A + B\eta} = \overline{A} - B\eta.$$

These patterns generalize to higher dimensional matrix algebra.

It is worth pointing out the first precursor to the quaternions: This precursor is the system

$$\{\pm 1, \pm\epsilon, \pm\eta, \pm i\}.$$

Here $\epsilon\epsilon = 1 = \eta\eta$ while $i = \epsilon\eta$ so that $ii = -1$. The basic operations in this algebra are those of epsilon and eta. Eta is the delay shift operator that reverses the components of the iterant. Epsilon negates one of the components, and leaves the order unchanged. The quaternions arise directly from these two operations once we construct an extra square root of minus one that commutes with them. Call this extra root of minus one $\sqrt{-1}$. Then the quaternions are generated by

$$\{i = \epsilon\eta, j = \sqrt{-1}\bar{\epsilon}, k = \sqrt{-1}\eta\}$$

with

$$i^2 = j^2 = k^2 = ijk = -1.$$

The "right" way to generate the quaternions is to start at the bottom iterant level with boolean values of 0 and 1 and the operation EXOR (exclusive or). Build iterants on this, and matrix algebra from these iterants. This gives the square root of negation. Now take pairs of values from this new algebra and build 2×2 matrices again. The coefficients include square roots of negation that commute with constructions at the next level and so quaternions appear in the third level of this hierarchy.

8.2 Matrix Algebra in General

Construction of matrix algebra in general proceeds as follows. Let M be an $n \times n$ matrix over a ring R. Let $M = (m_{ij})$ denote the matrix entries. Let π be an element of the symmetric group S_n so that $\pi_1, \pi_2, \cdots, \pi_n$ is a permutation of $1, 2, \cdots, n$. Let $v = (v_1, v_2, \cdots, v_n)$ denote a vector with these components. Let $\Delta(v)$ denote the diagonal matrix whose $i-th$ diagonal entry is v_i. Let $v^\pi = (v_{\pi_1}, \cdots, v_{\pi_n})$. Let $\Delta^\pi(v) = \Delta(v^\pi)$. Let Δ denote any diagonal matrix and Δ^π denote the corresponding permuted diagonal matrix as just described. Let $[\pi]$ denote the permutation matrix obtained by taking the $i-th$ row of $[\pi]$ to be the $\pi_i - th$ row of the identity matrix. Note that $[\pi]\Delta = \Delta^\pi[\pi]$. For each element π of S_n define the vector $v(M, \pi) = (m_{1\pi_1}, \cdots, m_{n\pi_n})$ and the diagonal matrix $\Delta[M]_\pi = \Delta(v(M, \pi))$.

Theorem. $M = (1/(n-1)!)\Sigma_{\pi \in S_n} \Delta[M]_\pi[\pi]$.

The proof of this theorem is omitted here. Note that the theorem expresses any square matrix as a sum of products of diagonal matrices and permutation matrices. Diagonal matrices add and multiply by adding and multiplying their corresponding entries. They are acted upon by permutations as described above. This means that any matrix algebra can be embedded in an algebra that has the structure of a group ring of the permutation group with coefficients Δ in an algebra (here the diagonal matrices) that are acted upon by the permutation group, and following the rule $[\pi]\Delta = \Delta^\pi[\pi]$. This is a full generalization of the case $n = 2$ described in the last section.

It is amusing to note that this theorem tells us that up to the factor of $1/(n-1)!$ a unitary matrix that has unit complex numbers as its entries is a sum of simpler unitary transformations factored into diagonal and permutation matrices. In quantum computing parlance, such a unitary matrix is a sum of products of phase gates and products of swap gates (forming the permutations).

A reason for discussing these formulations of matrix algebra in the present context is that one sees that matrix algebra is generated by the simple operations of juxtaposed addition and multiplication, and by the use of permutations as operators. These are unavoidable discrete elements, and so the operations of matrix algebra can be motivated on the basis of discrete physical ideas and non-commutativity. The richness of continuum formulations,

infinite matrix algebra, and symmetry grows naturally out of finite matrix algebra and hence out of the discrete.

9 Philosophical Appendix

The purpose of this appendix is to point to a way of thinking about the relationship of mathematics, physics, persons, and observations that underlies the approach taken in this paper. We began constructions motivating non-commutativity by considering sequences of actions $\cdots DCBA$ written from right to left so that they could be applied to an actant X in the order $\cdots DCBAX = \cdots (D(C(B(AX))))\cdots$. The sequence of events $A, B, C, D, \cdots$ was conceptualized as a temporal order, with the events themselves happening at levels or frames of successive "space". *There is no ambient coordinate space, nor is there any continuum of time.* All that is given is the possibility of structure at any given moment, and the possibility of distinguishing structures from one moment to the next. In this light the formula $DX = [X, J] = XJ - JX = J(X' - X)$ connotes a symbolic representation of the measurment of a difference across one time interval, nothing more. In other words DX represents a difference taken across a background difference (the time step). Once the pandora's box of measuring such differences has been opened, we are subject to the multiplicities of forms of difference $\nabla_K X = [X, K]$, their non-commutativity among themselves, the notion of a flat background that has the formal appearance of quantum mechanics, the emergence of abstract curvature and formal gauge fields. All this occurs in these calculi of differences *prior* to the emergence of differential geometry or topology or even the notion of linear superposition of states (so important to quantum mechanics). Note that in this algebraic patterning each algebra element X is an actant (can be acted upon) and an actor (via the operator ∇_X). In Lie algebras, this is the relationship between the algebra and its adjoint representation that makes each element of the algebra into a representor for that algebra by exactly the formula $adj_A(X) = [A, X] = -\nabla_A(X)$ that we have identified as a formal difference or derivative, a generator for a calculus of differences.

The precursor and conceptual background of our particular formalism is therefore the concept of discrimination, the idea of a distinction. A key

work in relation to that concept is the book "Laws of Form" by G. Spencer-Brown [41] in which is set out a calculus of distinction of maximal simplicity and generality. In that calculus a mark (denoted here by a bracket $< \ >$) represents a distinction and is seen to be a distinction between inside and outside. In this elemental mathematics there is no distinction except the one that we draw between the mathematician and the operator in the formal system as sign/symbol/interpretant. This gives full responsibility to the mathematician to draw the boundaries between the formal system as physical interaction and the formal system as symbolic entity and the formal system as Platonic conceptual form. In making a mathematics of distinction, the mathematician tells a story to himself/herself about the creation of a world. Spencer-Brown's iconic mathematics can be extended to contact any mathematics, and when this happens that mathematics is transformed into a personal creation of the mathematician who uses it. In a similar (but to a mathematician) darker way, the physicist is intimately bound to the physical reality that he studies.

We could have begun this paper with the the Spencer-Brown mark as bracket: $< \ >$. This empty bracket is seen to make a distinction between inside and outside. In order for that to occur the bracket has to become a process in the perception of someone. It has to leave whatever objective existence or potentiality it has alone (all one) and become the locus or nexus of an idea in a perceiving mind. As such it is stabilized by that perception/creation and becomes really a solution to $\{< \ >\} =< \ >$ where the curly bracket (the form of perception) is in the first place identical to the mark $< \ >$, and then distinguished from it by the act of distinguishing world and perceiver. It is within this cleft of the infinite recursive and the finite

$$< \ >= \{< \ >\} = \{\{< \ >\}\} = \{\{\{< \ >\}\}\} = \cdots = \{\{\{\{\{\cdots\}\}\}\}\}$$

that the objectivity of mathematics/physics (they are not different in the cleft) arises. All the rest of mathematics or calculus of brackets needs come forth for the observer in the same way. Through that interaction there is the possibility of a deep dialogue of many levels, a dialogue where it is seen that mathematics and physics develop in parallel, each describing the same boundary from opposite sides. That boundary is the imaginary boundary between the inner and outer worlds of an individual.

References

[1] M.F. Atiyah [1990], *The Geometry and Physics of Knots,* Cambridge University Press.

[2] Bastin, T. and Kilmister C. [1995], *Combinatorial Physics*, World Scientific Pub. Co.

[3] Bloom, Charles [1998], (private communication)

[4] Dyson, F. J. [1990], Feynman's proof of the Maxwell Equations, *Am. J. Phys.* 58 (3), March 1990, 209-211.

[5] R. Penrose [1971], Applications of negative dimensional tensors, In *Combinatorial Mathematics and Its Applications*, edited by D. J. A. Welsh, Academic Press.

[6] Crane, Louis , Kauffman, Louis H., Yetter, David N. [1997], State sum invariants of 4-manifolds, *Journal of Knot Theory and Its Ramifications*

[7] Connes,Alain [1990], *Non-commutative Geometry* Academic Press.

[8] Crane,Louis [1996], Clock and category: Is quantum gravity algebraic?, *J. Math. Phys.* **36** (11), November (1996), pp. 6180-6193.

[9] Crane, Louis [1997], A proposal for the quantum theory of gravity, arXiv:gr-qc/9704057 v2 23 Apr 97.

[10] Dimakis, A. and Müller-Hoissen [1992], F., Quantum mechanics on a lattice and q-deformations, *Phys. Lett.* 295B, p.242.

[11] Dirac, P.A.M. [1968], *Principles of Quantum Mechanics,* Oxford University Press.

[12] Etter, T. and Noyes, Pierre [2001], Process, System, Causality and Quantum Mechanics, *Bit String Physics*, edited by H. Pierre Noyes and J. C. van den Berg, World Scientific Pub. Co., 488-537.

[13] Forgy,Eric A. [2002] Differential geometry in computational electromagnetics, PhD Thesis, UIUC.

[14] E. Fredkin [1990], Digital Mechanics, *Physica D* **45**, pp. 254-270.

[15] Ashtekar,Abhay, Rovelli, Carlo and Smolin,Lee [1992], "Weaving a Classical Geometry with Quantum Threads", *Phys. Rev. Lett.*, vol. 69, p. 237.

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

[17] Hughes, R. J. [1992], On Feynman's proof of the Maxwell Equations, *Am. J. Phys.* 60, (4), April 1992, 301-306.

[18] Kauffman, L. [1985], Sign and Space, In Religious Experience and Scientific Paradigms. Proceedings of the 1982 IASWR Conference, Stony Brook, New York: Institute of Advanced Study of World Religions, (1985), 118-164.

[19] Kauffman, L. [1987], Self-reference and recursive forms, Journal of Social and Biological Structures (1987), 53-72.

[20] Kauffman, L. [1987], Special relativity and a calculus of distinctions. Proceedings of the 9th Annual Intl. Meeting of ANPA, Cambridge, England (1987). Pub. by ANPA West, pp. 290-311.

[21] Kauffman, L. [1987], Imaginary values in mathematical logic. Proceedings of the Seventeenth International Conference on Multiple Valued Logic, May 26-28 (1987), Boston MA, IEEE Computer Society Press, 282-289.

[22] Kauffman,Louis H.[1991,1994], *Knots and Physics,* World Scientific Pub.

[23] L. H. Kauffman, Knot Logic, In *Knots and Applications* ed. by L. Kauffman, World Scientific Pub. Co., (1994), 1-110.

[24] Kauffman,Louis H. and Noyes,H. Pierre [1996], Discrete Physics and the Derivation of Electromagnetism from the formalism of Quantum Mechanics, *Proc. of the Royal Soc. Lond. A*, **452**, pp. 81-95.

[25] Kauffman,Louis H. and Noyes,H. Pierre [1996], Discrete Physics and the Dirac Equation, *Physics Letters A*, 218 ,pp. 139-146.

[26] Kauffman,Louis H. and Noyes,H.Pierre (In preparation)

[27] Kauffman, Louis H.[1996], Quantum electrodynamic birdtracks, *Twistor Newsletter Number 41*

[28] Kauffman, Louis H. [1998], Noncommutativity and discrete physics, *Physica D* 120 (1998), 125-138.

[29] Kauffman, Louis H. [1998], Space and time in discrete physics, *Intl. J. Gen. Syst.* Vol. 27, Nos. 1-3, 241-273.

[30] Kauffman, Louis H. [1999], A non-commutative approach to discrete physics, in *Aspects II - Proceedings of ANPA 20*, 215-238.

[31] Kauffman, Louis H. [2002], Biologic. *AMS Contemporary Mathematics Series*, Vol. 304, (2002), pp. 313 - 340.

[32] Kauffman, Louis H. [2002], Time imaginary value, paradox sign and space, in *Computing Anticipatory Systems, CASYS - Fifth International Conference*, Liege, Belgium (2001) ed. by Daniel Dubois, AIP Conference Proceedings Volume 627 (2002).

[33] Manthey, M. [1999], A combinatorial bit-bang leading to the quaternions, Aspects II - Proceedings of ANPA 20, ed. by Keith Bowden, 23-45.

[34] Montesinos, M. and Perez-Lorenzana, A., [1999], Minimal coupling and Feynman's proof, arXiv:quant-phy/9810088 v2 17 Sep 1999.

[35] Müller-Hoissen,Folkert [1998], Introduction to non-commutative geometry of commutative algebras and applications in physics, in *Proceedings of the 2nd Mexican School on Gravitation and Mathematical Physics*, Kostanz (1998) ¡http://kaluza.physik.uni-konstanz.de/2MS/mh/mh.html¿.

[36] Noyes, Pierre [2001], *Bit String Physics*, edited by H. Pierre Noyes and J. C. van den Berg, World Scientific Pub. Co.

[37] Noyes, H. P., Kauffman, L.H., Lindsay, J.V., Lamb, W.R. [2003] , On E.D. Jones' microcosmology, SLAC-PUB-9620, Jan. 2003, arXiv:astro-ph/0301176 v1 10 Jan 2003.

[38] Penrose,Roger [1971], Angular Momentum - An Approach to Combinatorial Spacetime, In *Quantum Theory and Beyond*,Edited by Ted Bastin Cambridge University Press, pp. 151-180.

[39] Schweber, Silvan S. [1986], Feynman and the visualization of space-time processes, *Rev. Mod. Phys.* Vol. 58, No. 2, April 1986, 449 - 508.

[40] Smith, T., [1997] From sets to quarks, hep-ph/9708379 17 Aug 1997 via http://www.innerx.net/personal/tsmith/TShome.html

[41] G. Spencer-Brown, *Laws of Form,* Julian Press, New York (1969).

[42] Tanimura,Shogo [1992], Relativistic generalization and extension to the non-Abelian gauge theory of Feynman's proof of the Maxwell equations, *Annals of Physics, vol. 220*, pp. 229-247.

[43] Wene, G. P. [1982], A little color in abstract algebra,*Amer. Math. Monthly*, Vol. 89 No. 6 (June - July 1982), 417-419.

[44] Weyl, Hermann [1922], *Space-Time-Matter*, Methuen, London (1922).

[45] Witten,Edward [1989], Quantum field Theory and the Jones Polynomial, *Commun. Math. Phys.*,vol. 121,pp. 351-399.

On E.D.Jones' MICROCOSMOLOGY*

H. Pierre Noyes, Stanford Linear Accelerator Center, Stanford University
noyes@slac.stanford.edu
Louis H. Kauffman, Department of Mathematics, University of Illinois at Chicago,
kauffman@uic.edu
James V. Lindesay, Physics Department, Howard University,
jlslac@slac.stanford.edu
Walter R. Lamb, Energy Systems - Solar Inc.
148 Jacinto Way, Sunnyvale, CA 94086, LWrlamb@cs.com

Abstract

By taking seriously the limits on observability which come from combining relativistic quantum mechanics with general relativity, Ed Jones has shown that the current measurements of the cosmological constant density $\Omega_\Lambda \sim 0.7$ imply that the temperature scale at which it becomes possible to discuss cosmological models is $\sim 5\ Tev$ $(5.8 \times 10^{16}\ ^oK)$. This is self-consistent with the assumption that the number of Planck masses which make some sort of "phase transition" to this state is $N_{Pk} \sim 4 \times 10^{61}$. We review Jones' argument and the *bit-string physics* calculation which gives the baryon-photon ratio at nucleosynthesis as $\sim 2/256^4$, the dark matter-baryon ratio as ~ 12.7, and hence $\Omega_m \sim 0.3$, in agreement with current observations. Accepting these values for the two energy densities $\Omega_\Lambda + \Omega_m \sim 1$ in accord with recent analyses of fluctuations in the CMB showing that space is flat to about 6%. We conclude that experiments with particle accelerators in the 5-10 Tev range must either show that current theory can adequately describe the currently observed structure of our universe or force us to revise our ideas about physics at a very fundamental level.

*Work supported by Department of Energy contract DE–AC03–76SF00515.

I. INTRODUCTION

It has been known for some time that standard general relativistic cosmology can describe current astronomical observations in remarkable detail.[25, 26] Retrodicting from these results to earlier times, one finds extremely hot and dense radiation and matter ~ 13 *Gyr* ago. Conditions were such that no current experimental or astronomical systems are available to test what the "laws of physics" were then, or even whether such laws "existed". As E.D.Jones realized some time ago, one way to make progress in gaining a limited understanding of this "pre-physics" era is to take seriously the limitations that the shortest measurable length and time, and the largest measurable elementary mass and temperature place on physical cosmology. In this paper we review and discuss the remarkable discoveries that he made. We emphasize their fundamental character and independence of special assumptions.

E.D.Jones[10] discovered that a *neo-operational* [19] approach imposes fundamental limits on the measurement of short distances in such a way that we can *predict* that there must be a *positive* cosmological constant. Further, his preliminary calculation gave the cosmological constant energy density, relative to the critical density, of $\Omega_\Lambda = 0.6 \pm 0.1$. The basic idea is that we must find some way to connect the inchoate, pre-geometrical, pre-physical state of the universe by a "phase transition" to a state in which contemporary physics — explicitly, relativistic quantum mechanics and general relativity — can be consistently employed. The connection will be made by using one scaling parameter. This scaling parameter is fixed by requiring energy density equilibrium between the virtual energy density which thermalizes at some mass scale m and the virtual energy density $\rho_\epsilon = \frac{3\epsilon}{4\pi r_\epsilon^3}$ which is "left behind" when the (extremely rapid) transition to ordinary space, time and particles is complete. Here $r_\epsilon = \hbar c/\epsilon$. Since this residual virtual energy ϵ necessarily decouples from other contributions to the FRW equations for the evolution of the universe, Jones' interpretative postulate is to identify it with the cosmological constant density, i.e. $\rho_\epsilon = \rho_\Lambda$.

The limiting concepts we apply to obtain limits on observability are the uncertainty principle, the black hole surface area, and the cosmological event horizon. We make contact with experimental practice by assuming that we know four basic

universal constants in standard units. These are:

(1) Newton's gravitational constant $\mathbf{G_N}$, which can be fixed if we know the escape velocity of a particle from a system of radius R, and mass m, i.e.

$$\frac{\mathbf{G_N} m}{R} = \frac{1}{2} v_{escape}^2 \tag{1}$$

(2) Boltzman's constant $\mathbf{k}$ from the basic law of statistical mechanics connecting entropy, S, to the number of degrees of freedom W, i.e.

$$S = \mathbf{k}\ ln\ W \tag{2}$$

(3) Planck's constant $\mathbf{h}$ (or $\hbar = h/2\pi$), as in

$$E = \mathbf{h}\ \nu \tag{3}$$

(4) Einstein's limiting velocity $\mathbf{c}$, as in the mass-energy relation

$$E = m\ \mathbf{c}^2 \tag{4}$$

II. Shortest length; largest density; highest temperature

If there is a largest *elementary* fermion mass m_X, the shortest geometrical length in (special) relativistic quantum mechanics to which we can examine its (or any) structure is the Compton radius

$$R_Q(m_X) = \frac{\hbar}{2m_X c} \tag{5}$$

Wick pointed out long ago[27] that if we try to measure the structure of a system which couples to some mass μ, combining special relativity with the quantum energy-time uncertainty principle tells us that at distances less than $\hbar/\mu c$, which requires us to use available energies greater than μc^2, such a particle can be produced with a finite probability. For particles of mass m_X that satisfy the CPT theorem — no exceptions are known — and have a conserved quantum number which allows anti-particles to be defined (which includes all known fermions), examining the structure of m_X with *any* probe of energy greater than $2m_X c^2$ will necessarily produce a pair consisting of m_X and its anti-particle with some finite probability. But the m_X produced will be

indistinguishable from the m_X probed, and can appear *anywhere* within a distance $\hbar/2m_Xc$ of the target particle. Thus, in relativistic quantum mechanics (in contrast to non-relativistic quantum mechanics which in principle allows any short distance to be examined by using probes of high enough momenta) there is an absolute shortest distance, set by the largest elementary mass. Below this limit the concepts of point geometry simply *dissolve.*

In contrast, general relativity *starts* with geometrical concepts and would seem to be able (until we come to cosmology) to describe systems with arbitrarily large masses. However, if the mass is concentrated within a volume whose surface area is (for spherical symmetry) $4\pi R_G(m)^2$, we must use more care. "Inside" the *black hole* defined by this area and calculable from the Schwarzschild radius

$$R_G(m) = \frac{2mG_N}{c^2} \tag{6}$$

where G_N is Newton's gravitational constant, there is no known way to *measure* geometrical structure. Only the mass, and (if it has them) charge and magnetic moment can be measured by means of *exterior* observations. Consequently, in regions where the mass concentrations are small enough, we can use (if available) systems with small enough mass-energies to probe geometrical structure down to arbitrarily short lengths. Thus, by itself, general relativity seems to be able to use geometrical concepts consistently in any regions *outside* of black holes.

But any theory of *quantum gravity* must, at least in a limiting sense, combine both relativistic quantum mechanics and general relativity in a single theory. Since the shortest length observable in the first theory is inversely proportional to the mass considered, while the shortest length observable in the second theory is proportional to the mass considered, this requirement tells us (in the absence of new concepts) that there must be a finite limiting length for *any* theory of quantum gravity. We define this by setting the Schwarzschild radius equal to the Compton radius, i.e.

$$R_{min}^{QG} = R_G(m) = R_Q(m) \tag{7}$$

This leads immediately, to the requirement that $4m^2 = \hbar c/G_N \equiv M_P^2$.

The limiting mass M_P calculated in this way was first arrived at by Planck (using h), and is now (using $\hbar$) called the *Planck mass.* Planck's original argument was simply that G_N, k, c, and h were clearly universal constants. Since their dimensions in terms of (sometimes fractional) powers of mass, length, time, and temperature are *independent,* these constants collectively define what are certainly universal, and perhaps fundamental dimensional units for physics. But the numerical values so predicted were discouraging. The *Planck mass* in SI units is approximately 2.18×10^{-8} kilograms. Even today there is no known elementary particle with this mass. The corresponding *Planck length* $L_P \equiv \hbar/M_P c = 1.62 \times 10^{-35}$ *m*, *Planck time* $T_P \equiv \hbar/M_P c = 5.39 \times 10^{-44}$ *sec* and *Planck temperature* $\Theta_P \equiv M_P c^2/k = 1.42 \times 10^{32}\ ^oK$. These sounded ridiculous if one wished to relate them to available methods of measuring masses, lengths, times and temperatures. But as *limiting* quantities, they make sense. To our knowledge, Jones' simple calculation, which at least for us makes this interpretation compelling, originated with him.

That it makes no sense using current theory to talk of lengths less than the Planck length (or times less than the Planck time, or elementary masses greater than the Planck mass, or temperatures higher than the Planck temperature) is hardly a novel conclusion. But Ed Jones' demonstration of the fact, which we have just summarized, is so straightforward that we think it deserves wide circulation.

What is, perhaps, less familiar is that the same physical principles prohibit meaningful discussion of physical matter at any density greater than the Planck density, i.e. a Planck mass in a Planck volume which must exist for at least a Planck time, that is for at least $[\frac{\hbar G_N}{c^5}]^{\frac{1}{2}}$. We define this limiting density scale by

$$\rho_P \equiv \frac{M_P}{\frac{4\pi}{3}L_P^3} = \frac{3}{4\pi}[c/\hbar]^3 M_P^4 \tag{8}$$

From now on we will use units in which $\hbar = 1 = c = 1 = k = 1$, so that any mass μ can stand (ambiguously) for either a mass or an energy or a temperature; any inverse mass μ^{-1} can stand (ambiguously) for a length or a time. We see that in these units $\frac{3}{4\pi}\mu^4$ is the energy density of mass-energy μc^2 in a spherical volume whose radius is its Compton wavelength.

We now show that in these units $\frac{3}{4\pi}M_P^4$ is, indeed, the limiting energy density.

Consider first masses less than a Planck mass inside a Planck volume. For such systems, the Compton wavelength exceeds the Planck length, and the limiting density is achieved only when the contained mass is equal to the Planck mass. Consider next systems with a mass greater than the Planck mass which we try to confine within a Planck volume. For such systems, the Schwarzschild radial coordinate defined by the horizon area exceeds the size of the system we are attempting to construct, and to assume that it is confined within a smaller radial coordinate is, again, meaningless. Although the radial coordinate does not measure proper distance, it nonetheless sets the radial scale of the geometry relative to the horizon. So only the limiting case of systems at the Planck density (or less) whose volume just exceeds the volume measured by the Schwarzschild radius (or is greater) have any hope of being given observational meaning. Since these two cases exhaust the possibilities, no system with greater than the Planck density can be given operational meaning. Q.E.D.

For completeness, we also note that the same considerations we used to establish limits on elementary masses, lengths and times also limit the maximum temperature to the Planck temperature. At first sight, this is not completely obvious. We are allowed massless radiation, and hence could define temperature by the Planck distribution in, for example, an Einstein-De Sitter universe that contains only gravitation and electromagnetic radiation. However radiation, by itself, cannot come to equilibrium even when the individual frequency modes are quantized using the $E = h\nu$ prescription for each degree of freedom. These modes could be initially set to any arbitrary distribution. Without some interaction between the modes this arbitrary initial distribution will never come to the Planck distribution. Hence, there is no way to define "temperature".

To obtain insight on how relativistic quantum mechanical temperature measurement necessarily gets connected to a mass scale it is useful to go back to the classic paper by Bohr and Rosenfeld[4] which provided the initial operational justification for "second quantization" of the electromagnetic field. They derived the commutation relations between electric and magnetic field strengths by applying the uncertainty principle to the material apparatus which measures them. This derivation can avoid any discussion of the particulate structure of matter because the only universal con-

stants they invoke are $\hbar$ and c. Thus the theory they discuss is scale invariant and their analysis need only use wavelengths so long that the classical description of matter can be invoked consistently. However, as they point put, when wavelengths of the order of $\hbar/mc$ are required, which is true at mass scale m, their derivation breaks down. We then need to use the coupling between the matter field and the electromagnetic field in order to study the meaning of measurement. This, in turn, brings in the scattering of light by light due to the creation (real or virtual) of particle-antiparticle pairs. This coupling then provides transitions between electromagnetic modes and in the black body situation leads, sooner or later, to the Planck distribution. Clearly this same mechanism can achieve thermalization in the cosmological context. Once more this ties us to some finite mass scale, and the arguments already given limit us to temperatures below the Planck temperature.

Since any system we consider, including our observed universe, cannot be described using current physics if it exceeds the Planck density, we next ask whether assigning the Planck density to our universe at some stage in its history makes sense. Consider first the case when the universe that starts at this density has one Planck mass. But then its event horizon, or radial scale factor in the FRW metric, coincides with the Schwarzschild radius for that mass. There is no space "outside" this radius which can allow us to define the volume which this mass occupies, so the concept of density remains vacuous. However, if this universe has a larger radial parameter (and hence a lower density) allowing it to become a "black hole" surrounded by an "ergo-sphere" made up of particles and anti-particles and radiation at some mass-energy-temperature scale $m < M_P$, it becomes a possible starting point for a universe which becomes describable at an event horizon parameterized by $R_H \sim 1/m$. Jones realized that these arguments are closely related to the Dyson-Noyes argument[18, 20], as we now show.

Dyson[6] pointed out that if there are $Z_{e^2} = \alpha_{e^2}^{-1} \simeq 137$ electromagnetic interactions within the Compton wavelength of a single charged particle-antiparticle pair (i.e. $\hbar/2mc$), there is enough energy to create another pair. Whether these interactions are virtual, or real (eg in a system with enough energy and appropriate internal momenta to concentrate $2mc^2 Z_{e^2}$ of that energy within this Compton wavelength), in

a theory for which like charges attract rather than repel each other still more energy can then be gained by creating another pair; the system collapses to negatively infinite energy. Dyson concluded that the renormalized perturbation theory for QED is not uniformly convergent beyond 137 terms. Note that this bound can be written as $Z_{e^2}\alpha_{e^2} = 1$. Noyes[18, 20] noted that for electron-positron pairs, this critical energy corresponds approximately to the threshold for producing a pion. This fact provides a physical interpretation of the reason for the failure of QED: QED ignores strong interactions mediated by pions, or more generally by quarks and anti-quarks which bind to yield pions.

For gravitation the corresponding coupling constant $\alpha_m = G_N m^2 = m^2/M_P^2$ and the critical condition becomes

$$Z_m \alpha_m = 1 \; or \; Z_m = \frac{M_P^2}{m^2} \tag{9}$$

where Z_m represents the number of gravitational interactions within $\hbar/mc$ defining this critical condition. That is, for quantum gravitational perturbation theory, the cutoff mass-energy corresponds to the Planck mass rather than the pion mass, which makes sense.

III. Plancktons: The Pre-physics - Physics Transition

To see how a dense system like that discussed at the end of the last section might make sense in a cosmological context, consider first a system which starts from a Planck's mass worth of quantized mass-energy (at some mass scale m where we have confidence that currently accepted physics is applicable) distributed in a radial shell with radially inward-directed momentum. Such a system, assuming no changes in the physics along the way, would have a finite chance of making it down to the Planck density scale before rebounding (or whatever). This suggests that we might, in some sense, be able to to describe a universe which "starts" with one *Planckton* (which we define as a Planck mass in a Planck volume and hence necessarily at the Planck temperature). If our current physics is not capable of describing a single Planckton, it is still possible to envisage some non-adiabatic expansion process which would allow the virtual energy "contained" in this Planckton to be distributed throughout

a large volume at some small enough mass scale and event horizon large enough so that *after this expansion* this universe *will be describable* using currently understood physics. This assumes, of course, that the mass scale m is low enough that we have confidence that current theory can describe Z_m gravitational interactions within volumes contained within the event horizon whose linear measure is $\sim 1/m$, and that the Dyson-Noyes analysis applies.

Of course the universe described in the last paragraph is not *our* universe, which is known to contain (and have contained for twelve thousand million years or more) a lot more than one Planckton's worth of mass. However, since we are postulating a non-adiabatic transition from the Planck density to situations we *can* describe, it makes just as much sense to start from N_{Pk} Plancktons as to start from one. Of course this whole assemblage would have to be at the Planck density or just below it; on another occasion we hope to be able to discuss whether such an assemlage can be consistently described as a state of *Plancktonic matter.* This idea allows us to think of the initial expansion as starting from a virtual energy state and making a transition to a real energy state at some mass scale m, *leaving some of that virtual energy behind.*

Assume that the event horizon at which this residual virtual energy ϵ has observable consequences can be characterized by a radial parameter which we call $R_H(\epsilon) \sim 1/\epsilon$. Further assume that the transition point corresponds to some expansion factor Z_ϵ from the Planck length, i.e.

$$R_H(\epsilon) \equiv Z_\epsilon L_P = \frac{Z_\epsilon}{M_P} \sim \frac{1}{\epsilon} \tag{10}$$

Here ϵ is the energy *per Planckton.* One might note that this defines the expansion factor in terms of R_H, which will give a measure of the square root of the area of the horizon, NOT a radial distance. This is a key feature of holographic interpretations of the physical properties of horizons[3]. This relationship also associates the energy per Planckton with the expansion scale of this horizon from the Planck length.

The corresponding energy density scale which this residual virtual energy has at the moment when the phase transition is complete is $\rho_\epsilon = N_{Pk}\frac{3}{4\pi}\epsilon^4$. Since this came from N_{Pk} Plancktons, each confined within a volume scale given by a single Planckton,

we can also write down the event horizon (scale factor, "Schwarzschild radius of the universe") $R_H(\epsilon)$ defined by this mass scale as

$$R_H(\epsilon) = G_N N_{Pk}\epsilon \Rightarrow 1 \sim \frac{N_{Pk}\epsilon^2}{M_P^2} = \frac{N_{Pk}M_p^2}{M_P^2 Z_\epsilon^2} \Rightarrow N_{Pk} = Z_\epsilon^2 \tag{11}$$

Jones now assumes that the bulk of the energy which started out as N_{Pk} Planckons thermalizes at mass scale m and that the corresponding energy density $\rho_m = \frac{3}{4\pi}m^4$ is in energy equilibrium with the residual energy ρ_ϵ at the moment of completion of the phase transition. Hence

$$N_{Pk}\epsilon^4 = m^4 \Rightarrow \frac{N_{Pk}}{Z_\epsilon^4} = \frac{m^4}{M_P^4} = \frac{1}{Z_m^2} \Rightarrow Z_m^2 = Z_\epsilon^2 \tag{12}$$

Thus the Dyson-Noyes factor Z_m (which allows for operationally definable coordinates in the quantum domain) coincides with the residual virtual energy expansion scale factor Z_ϵ. This is true only at the completion of the non-adiabatic phase transition (at which point there are also operationally definable coordinates in the gravitational/geometrical domain). Hence we conclude that

$$Z_\epsilon = Z_m = Z = N_{Pk}^{\frac{1}{2}} \quad and \quad m^2 = \epsilon M_P \tag{13}$$

Jones also points out that this fact is over-constrained because Z_ϵ and Z_m each corresponds to the entropy (number of degrees of freedom) of each of the systems, which must also be equal at equilibrium. We also note that the mass scale at which the transition becomes complete is the geometric mean between the residual virtual energy and the Planck mass. Since the system starts in the pre-physics regime in which geometric structure cannot be specified, in the resulting thermalized state, prior to further evolution, "where" in the earlier state any Planckton's worth of energy "came from" also cannot be specified. Hence the virtual energy which is momentarily in density equilibrium with the mass-energy at mass scale m is *uniformly* distributed. Consequently the universe describable using current physics *starts out* with no structure even though it is much too large for the regions to be causally connected. Thus, Jones' use of fundamental physical principles solves the "horizon problem" *without* having to postulate the unknown "physics" implicit in the currently popular "inflationary" scenarios.

We now complete the Jones argument. Note that since the residual energy density must be positive, and since the transition — whatever the details — must be extremely rapid, it can be seen that ρ_ϵ corresponds to the cosmological constant density "boundary condition" ρ_Λ in the FRW equations. That it is positive is required for logical consistency, because this corresponds to a "negative pressure" (expansive force)[8] which makes the transition *irreversible.*

We can now check this conclusion by comparison with cosmological observations. Putting together the algebraic results already established and the definition of the cosmological constant energy density in terms of the critical density ($\rho_\Lambda = \Omega_\Lambda \rho_c = \Omega_\Lambda \rho_\epsilon$), we have that the basic scale parameter

$$Z = (\frac{\rho_P}{\Omega_\Lambda \rho_c})^{\frac{1}{4}} = (\frac{0.7}{\Omega_\Lambda})^{\frac{1}{4}} (\frac{0.71}{h_0})^{\frac{1}{2}} \times 6.564 \times 10^{30} \tag{14}$$

For the Planck density we use Eq. 8, which works out to be $\rho_P = 6.906 \times 10^{117} \frac{Gev/c^2}{cm^3}$. For the critical energy density we use[25] $\rho_c = 1.054 \times 10^{-5} h_0^2\ Gev/c^2\ cm^{-3}$. Finally we accept $\Omega_\Lambda = 0.7$ for the normalized cosmological constant density and $h_0 = 0.71$ for the normalized Hubble constant, as is indicated in Eq. 14. It is then easy to calculate the thermalization mass scale from the Dyson-Noyes relation (Eq. 9) as

$$m = Z^{-\frac{1}{2}} M_P = (\frac{\Omega_\Lambda}{0.7})^{\frac{1}{8}} (\frac{h_0}{0.71})^{\frac{1}{4}} \times 4.766\ Tev/c^2 \tag{15}$$

The positive cosmological constant, let alone its value, was still a matter of debate two or three years ago, resting as it did solely on the measured luminosity and red shifts of a number of distant type Ia supernovae (IaSne). These results have recently been improved[26]. As was pointed out this spring by Frieman[9], lingering doubts can be set to rest by the fact that a completely different type of evidence now shows that $\Omega_\Lambda = 0.7$, consistent with the Type Ia supernova data. The new evidence is simply that *fluctuations* in the cosmic microwave background radiation show that our universe is *flat* to about 6%, i.e. that $\Omega_m + \Omega_\Lambda = 1$ where Ω_m is the normalized mass-energy density. Since it is known from a number of different types of data that $\Omega_m = 0.3$, the value $\Omega_\Lambda = 0.7$ follows immediately.

We now assert that Jones has *proved* his contention that $\Omega_\Lambda = 0.7$ implies a mass scale of $\sim 5\ Tev/c^2$ ($(5.8 \times 10^{16}\ ^oK)$ or visa versa. To our knowledge, this is the

first time that the Planck mass has been directly and *quantitatively* connected to any observable phenomenon. We emphasize that although the Jones argument connects a conventional FRW universe to a denser situation where conventional concepts loose operational meaning, *all* that he requires is that normal scaling holds across the phase transition, and depends on only one scale factor. That there is only one independent scale factor is the conclusion acceptance of Occam's razor would establish directly.

We also wish to emphasize that our reproduction of E.D.Jones calculation here makes no claim to novelty, and is presented prior to the posting of his own paper because of an unexpected delay in the presentation of his own way of looking at the problem. We stress that his thinking contains novel elements not discussed here and that our cruder discussion should not be used as a substitute for his work as soon as that becomes available.

IV. Relation to Bit-String Physics

We have seen that Jones' calculation depends on one fundamental dimensionless scaling parameter, the number of Plancktons ($N_{Pk} = Z^2 \sim 4.3 \times 10^{61}$) which thermalize at mass scale $m \sim 5\ Tev$ leaving behind energy density ϵ per Planckton. Since $m^2 = \epsilon M_P$ (Eq. 12), any fundamental, dimensionless theory which allows us to (a) identify within its structure the Planck mass and three other dimensionless structural constants which bear a known, and mutually independent, connection to, e.g. c, $\hbar$ and k and (b) calculate Z (or m or ϵ) would allow us to say we have a first order, fundamental understanding of physical cosmology. Where could we find or how could we construct such a theory?

The candidate theory we examine here is bit-string physics[23]. We choose this construction both because Jones' theory is historically connected to the research program that led to bit-string physics (specifically, by his use of the Dyson-Noyes argument), and because the only places where his prediction that $\Omega_\Lambda \sim 0.6$ appear are in papers stimulated by his private communication of that result to HPN[21, 22]. Bit-string physics in turn arose out of the combinatorial hierarchy of Amson, Bastin, Kilmister and Parker-Rhodes[24], which in turn came out Bastin and Kilmister's[2] interest in Eddington's search for a fundamental theory[11]. One reason this research

program looks promising is, among other things, because of Eddington's contention that dimensionless numbers like $\alpha_{e^2} = \hbar c/e^2 \approx 137$ get into physics only when we start from a logical or mathematical "pre-physics" and/or "pre-geometry" construction. Then such dimensionless numbers could arise *prior* to development of procedures we can relate to conventional measurement and which give quantitative meaning to dimensional symbols like $\hbar, c$ and e^2 in any consistent system of physical units. To one of us (HPN), the program universe construction which was created as part of the research program on the combinatorial hierarchy sounds suspiciously like Jones' postulated "non-adiabatic process" which takes the universe from the Planck scale up to a scale where current physics clearly applies.

Unfortunately, none of this fundamental work on the combinatorial hierarchy or bit-string physics has as yet led to a *reliable, quantitative* connection to mainstream physics. Critically viewed, one might say that this work consists of a few numerical coincidences which are still in search of what might deserve to be called a speculative physical theory. We contrast this with Jones' theory, which we hope the first three sections have shown *does* constitute a theory based on a few fundamental and generally accepted principles in mainstream physics. Despite these critical remarks, we include a brief discussion of the connection to bit-string physics here in the hope that this discussion may help in planning future research.

We start with the best succinct, accurate and published statement of what the *combinatorial hierarchy* is. This is due to McGoveran[13]:

> The Combinatorial Hierarchy is generated from two recursively generated sequences. The first is governed by the recursion formula $n_{i+1} = 2^{n_i} - 1$ (a formula familiar to those who have studied the Mersenne primes), and begins with the term n=2 leading to the sequence $3, 7, 127, 2^{127} - 1, ...$ The cumulative cardinals of this series (ignoring the initial term) also form a series which has interpretive significance, namely $2, 3, 10, 137, \sim 1.7016... \times 10^{38} + 137, ...$
>
> The second recursively generated sequence is governed by the formula $m_{i+1} = m_i^2$. These two sequences have various justifications. Perhaps the

clearest presentation has been given by Clive Kilmister (correspondence to H.P.Noyes date Oct. 16, 1978), paraphrased here as follows:

Definition:

By a combinatorial hierarchy is meant a collection of levels selected as follows:

a) the elements of level L are a basis of a vector space V/Z_2

b) the elements at Level L+1 are non-singular (i.e. invertible) linear operators mapping V/Z_2 into V/Z_2

c) each element A at level L+1 are mapped to a subset S of the elements at Level L by the correspondence: the proper eigenvalues of A [i.e, Av = v] are exactly the linear subspace generated by S.

d) each element at level L+1 is chosen independent, allowing the process to be repeated for level L+2, L+3, L+4, ...

Theorem 1:

There is a unique hierarchy (up to isomorphism) with more than 3-levels and it has the following successive numbers of elements: $2, 3, 7, 127, 2^{127}-1$ and terminates at level 4 due to the fact that the operators have m^2 elements if the vectors are m-fold and 2^n (required for V/Z_2) increases too fast.

What strikes some of us when we encounter the cumulative sequence is that the third term $137 \approx \hbar c/e^2 = \alpha_{e^2}$ is the inverse fine structure while the number of elements in the terminating level $2^{127} + 136 \approx 1.7016 \times 10^{38} \approx \hbar c/G_N m^2_{proton} = M_P^2/m^2_{proton}$ is the square of the ratio of the Planck mass to the proton mass. Note that one is the Dyson-Noyes number characterizing electromagnetic interactions, and the other the corresponding Dyson-Noyes number for the gravitating particles which constitute most of the known paticulate mass of the universe (if the "dark matter" can be shown *not* to be particulate, or about 10% of the total mass of the universe if the "dark matter" *is* particulate). This looks promising for a cosmology based on this

construction. But progress toward a physical theory has been distressingly slow. A historical summary has been published[20]. Perhaps the most encouraging results are the derivation of the Sommerfeld formula and calculation of the next four significant figures (beyond 137) in the inverse fine structure constant by McGoveran[15, 17] and his various combinatorial corrections to other results, reprinted and discussed in[16]. All of these corrections improve the fit to experiment.

Despite the vagueness of the contact between bit-string physics and relativistic quantum particle dynamics, the structure already discussed suggests a way to calculate two cosmological parameters[21]. The first is the dark matter to ordinary matter ratio. In the hierarchy construction, we do not encounter the connection to electromagnetism until we have constructed level 3 (i.e. the level characterized by $137 = 127 + 10$ cumulative elements). However, one string in level 4 interacts gravitationally with everything. This strongly suggests that that the first two levels $10 = 3 + 7$ do not interact electromagnetically but do interact gravitationally, and could therefore be used to represent dark matter, whether it is particulate or non-particulate. Then, if we use a constructive algorithm with an arbitrary, stochastic element (as is done, for example, in *program universe* — see[20] for details), the usual assumption that in the absence of further information all elements receive equal weight immediately predicts that the dark matter to electromagnetically interacting matter ratio is $127/10 = 12.7$. At the time of nucleosynthesis, it is also plausible to assume that the the dark matter to electromagnetically interacting (at the fundamental level) matter, can be approximated by the dark matter to baryonic matter ratio, so we make that further assumption. The final step needed to connect to cosmological observation (see, e.g.[25, 26]) is the photon number to baryon number ratio at the time of nucleosynthesis.

To get the photon/baryon ratio, we need only slightly more detail about our constructive algorithm than has already been invoked. As is discussed in more detail in the paper already cited[21], the most likely bit-strings a stochastic construction yields, i.e. those (when of even length $2N$) with the number of zeros (N_0) equal to the number of ones (N_1) — hence $N_0 = N_1$, $N_0 + N_1 = 2N$, are also the prime candidates for representing photon labels (quantum numbers). The next most likely strings

have $N_0 = N_1 \pm 1$ or, with equal probability $N_0 = N_1 \mp 1$. Note that these strings with an odd number of ones, conserve this characteristic when "interacting" with the photon (even number of ones) strings as modeled by our basic operation of addition in Z_2. This in turn suggests a conserved quantum number such as baryon number. Further, the more detailed interpretation of the basic *progam universe* algorithm in the reference already cited[20] seems to yield the type of driving terms (two-body scatterings) needed for a finite particle number relativistic quantum mechanical scattering theory[1].

The basic processes needed to estimate the photon baryon ratio in the cosmological context of nucleosysnthesis we are discussing are then the probability of a photon-photon scattering process with a similar process as "spectator" (in the few body sense) compared to a photon-baryon scattering process with the same spectator. In the first case there are four photons in the initial (and final) state, and in the second there is one baryon and three photons in the initial (and final) state. At this stage in the construction (level 4 completed) the labels are strings of length 256. To change the photon label to a baryon label in one of the four photon labels, obtaining the next most probable process, can only happen in one out of 256^4 ways, giving a baryon-photon ratio of $1/256^4 \sim 2.3 \times 10^{-10}$. When first presented[21], this result was in comfortable agreement with what was then known about this number from cosmic abundances of the primordial nuclei, and predicted a value of Ω_m which was, perhaps, a little low, but amazingly good for such a speculative calculation.

Recent observations, particulary of the primordial deuterium to hydrogen ratio as inferred from the absorption spectra when the interstellar and intergalactic deuterium is illuminated by very early quasars, have tightened the allowed values of the baryon-photon ratio and moved the median up nearly a factor of two, as estimated by Fields and Sarkar[7]. Their recommended limits now just exclude HPN's value and suggest that there may have been an error in his reasoning. The error was that the $1/256^4$ value ignored the fact that *both* the case when N_1 exceeds N_0 by one *and* the case when it is one less should have been counted, providing the "missing" factor of two. If accepted, this correction has the added advantage of moving the prediction of Ω_m closer to the median of the observed value. See[21] for the details of the original

calculation. We conclude here that HPN's calculation does provide a weak indication that bit-strig physics predicts $\Omega_m \sim 0.3$, but that more work is badly needed before much confidence can be placed in it.

If one accepts the bit-string prediction that $\Omega_m \sim 0.3$ *and* the empirical fact[9] that space is flat to about 6%, we can then use the constraint for flat space $\Omega_\Lambda + \Omega_m = 1$ to say that bit string physics *predicts* $\Omega_\Lambda \sim 0.7$. Then, as we have seen in our discussion in the first three sessions Jones' MICROCOSMOLOGY shows that the dark energy density at the current time can be understood, that thermalization/baryon number conservation/physical space-time all became meaningful at a mass scale $m \sim 5Tev$, and all started from a pre-physics, pre-geometry universe characterized by a virtual energy of $N_{Pk} \sim 4 \times 10^{61}$ Planck masses.

Clearly this puts a high premium on tightening up the connection between bit-string physics and mainstream physics to the point where it can enter the field as a respectable contender for the research interest of the physics and astrophysics community. Where to begin is a matter of taste. One possibility, which HPN favors, is to try to show that dark matter is particulate with a unique mass of $m_D \sim 5\ Tev/c^2$ and *only* gravitational interactions. Searches for dark matter in this range are already under way; having a prediction to confirm or refute might help the experimenters. On the theoretical side, we note that bit-string physics can already claim to have provided a kind of strong-electromagnetic unification by arguing that the pion mass can be approximately calculated as twice the inverse fine structure constant in units of the electron mass. To get weak-gravitational unification, we might argue by analogy with the weak-electromagnetic unification (including the Coulomb force) which gives the heavy vector mesons (W,Z) and one or more massive Higgses, that electromagnetic-gravitation unification (including Newton's gravitational force) might give a massive scalar at $\sim 5\ Tev/c^2$. This is, admitttedly, a long shot, but might provide amusement for those with appropriate skills.

V. Conclusions

Accepting the growing consensus that current observations inescapably require that $\Omega_\Lambda = 0.7$ to 20% or better, Jones' reasoning allows him to use this value to state

that the mass-energy-temperature scale at which "pre-physics" makes a transition to observable space, time and particles is ~ 5 *Tev*! In other words, the next generation of particle accelerators will either substantiate the view that we *already* have in hand enough physics to understand the basic structure parameters of the currently observed universe *quantitatively*, or the next generation of high energy particle physics machines (if constructed) will necessarily provide graphic evidence that fundamental revision of our basic concepts is needed. We trust we have made it clear that bit-string physics has the basic logical structure needed to provide the one theoretical scaling factor on which Jones' theory rests. Optimists will go further and think that it already provides weak support for MICROCOSMOLOGY.

Acknowledgment

We are most grateful to E.D.Jones for permitting us to post this discussion of his theory before his own paper is readily available. We emphasize again that our paper should not be used as a substitute for his original work, particularly since we have used a number of short cuts and simplifications in our presentation, and tied it to bit-string physics in a way that has not been discussed with him.

References

[1] M.Alfred, P.Kwizera, J.V.Lindesay and H.P.Noyes, "A Non-Perturbative, Finite Particle Number Approach to Relativistic Scattering Theory, hep-th/0105241; submitted to *Foundatins of Physics*. This general theory of a non perturbative relativistic, finite particle number, quantum mechanics, has the appropriate particle-antiparticle symmetries and models with exact solutions of the quantum mechanical Coulomb and gravitational two-body problems in the appropriate limits[14], and gives the usual, manifestly covariant amplitude for the Compton scattering relevant here in the weak coupling limit[12].

[2] T.Bastin and C.W.Kilmister, *Combinatorial Physics*, World Scientific, Singapore, 1995; Vol. 9 in *Series on Knots and Everything*.

[3] D.Bigatti and L.Susskind, *TASI lectures on the Holographic Principle*, hep-th/0002044.

[4] N.Bohr and L. Rosenfeld, *Det. Kgl. Danske Videnskabernes Selskab., Mat.-fys. Med.* **XII**, (8).

[5] K.Bowden,Guest Editor, Physical Systems Special Issue, *Int.J.Gen.Systems*, Vol. 27, Nos, 1-3 (1998).

[6] F.J.Dyson, *Phys.Rev.*, **85**, 631 (1952), and a talk at about that date.

[7] B.D.Fields and S. Sarkar[26], Sec. 19, Fig. 19.1, p 162.

[8] Frieman[9] has noted that General Relativity *requires* that the equation of state for the "dark energy" currently observed have a negative pressure.

[9] "Cosmology", in *Particles in Collision*, conference at Stanford, paper given at 9:00 AM on June 21, 2002.

[10] E.D.Jones, private communication to HPN c. 1997. The result of Jones' calculation which gave $\Omega_\Lambda = 0.6 \pm 0.1$ is quoted in *Aspects II (Proc. ANPA 20)*, K.G.Bowden, ed. (1999), p. 207, and cited again in[23], p. 561.

[11] C.W.Kilmister, *Eddington's Search for a Fundamental Theory*, Cambridge, 1994.

[12] J.V.Lindesay and H.P.Noyes, "Non-Perturbative, Unitary Quantum-Particle Scattering Amplitudes from Three-Particle Equations", hep-th/0203262.

[13] David McGoveran, in[15], pp 38-39.

[14] J.V.Lindesay and H.P.Noyes, "Construction of Non-Perturbative, Unitary Particle-Antiparticle Amplitudes for Finite Particle Number Scattering Formalisms", nucl-th/0203042.

[15] David McGoveran, "The Fine Structure of Hydrogen", in *Discrete Physics and Beyond, Proc. ANPA 10*, C.W,Kilmister, Ed., 1988, pp 28-52.

[16] David McGoveran, "Comment on 'Our Joint Work ' ", in[23], pp 297-317.

[17] D.O,McGoveran and H.P.Noyes, *Physics Essays*, **4**, 115-120 (1991); reprinted as Ch. 6 in[23], pp 291-296.

[18] H.P.Noyes, "Non-Locality in Particle Physics" prepared for an unpublished volume entitled *Revisionary Philosophy and Science*, R.Sheldrake and D.Emmet, eds; available as SLAC-PUB-1405 (Revised Nov. 1975); reprinted as Ch. 1 in [23].

[19] H.P.Noyes, "Operationalism Revisited: Measurement Accuracy, Scale Invariance and the Combinatorial Hierarchy", *Science Philosophy Interface*, **1**, (1996), pp 54-79; reprinted as Ch. 10 in[23].

[20] H.P.Noyes "A Short Introduction to BIT-STRING PHYSICS", *Proc. ANPA 18*, reprinted in [23], as Ch. 14, esp. Sec. 1.1, pp 447-49 and Sec. 12, pp 478-481.

[21] H.P.Noyes, "Bit-String Physics Prediction of η, the Dark Matter/Baryon Ratio and Ω_M, Contributed paper presented at DM2000; reprinted as Ch. 16 in [23]. A shortened version appears in [22].

[22] H.P.Noyes, "Prediction of Ω_M and Estimate of Ω_Λ", in *Sources and Detection of Dark Matter and Dark Energy in the Universe*, D.B.Cline, ed., Springer (2000), pp 275-280.

[23] H.P.Noyes, *Bit-String Physics*, World Scientific, Singapore 2001, Vol. 27 in *Series on Knots and Everything.*

[24] A.F.Parker-Rhodes and J.C.Amson, "Hierarchies of Descriptive levels in Physical Theory", in [5], pp 57-80.

[25] Particle Data Group, **Astrophysics and cosmology**, *Euro. Phys.J.* **C 15**, (2000), p. 74.

[26] Particle Data Group, **Astrophysics and cosmology**, *Phys.Rev.* **D 66**, 148-188 (2002).

[27] G.C.Wick, *Nature*, **142**. 993 (1938).

Algebraic Quantum Mechanics, Algebraic Spinors and Hilbert Space.

B. J. Hiley.

Theoretical Physics Research Unit, Birkbeck, Malet Street, London WC1E 7HX.

b.hiley@bbk.ac.uk

Abstract.

The orthogonal Clifford algebra and the generalised Clifford algebra, C^n, (discrete Weyl algebra) is re-examined and it is shown that the quantum mechanical wave function (element of left ideal), density operator (element of a two sided ideal) and mean values (algebraic trace) can be constructed from entirely within the algebra. No appeal to Hilbert space is necessary. We show how the GNS construction can be obtained from within both algebras. The limit of C^n as $n \rightarrow \infty$ is shown to be the extended Heisenberg algebra. Finally the relationship to the usual Hilbert space approach is discussed.

1. Introduction.

In my study of the structure of Clifford algebras particularly in regard to the role they play in quantum mechanics through Pauli spinors, the Dirac spinors and Penrose's twistors, I was struck by how all the relevant results could be accounted for from within the algebra. There was no need to leave the algebra and to construct an external Hilbert space upon which the elements of the algebra were deemed to act when they play the role of observables. Of course all of these algebras are finite and non-nilpotent so that they are easily treated by standard algebraic analysis. My original thoughts on these topics can be found in Frescura and Hiley (1980, 1984 and 1987).

However when it comes to the Heisenberg algebra these techniques appear not to be applicable because this algebra is nilpotent. However the earlier work of Schönberg (1957) and the later ideas of Frescura and Hiley (1989) and Hiley (1991) suggested there was such a possibility provided one extended the algebra (Hiley 2001). Considerable

clarification of the meaning of this specific extension has been achieved by exploring the finite discrete Weyl algebra[1] C^n. (Morris 1967 and Hiley and Monk 1993). The importance of these algebras for topic of discussion in this paper is that in the limit as $n \rightarrow \infty$ the algebra contains the Heisenberg algebra. These algebras are finite so that the same techniques used in Clifford algebras can be directly applied. Thus we have a unified way of dealing with the basic algebras that lie at the heart of quantum mechanics.

These results have only intensified my curiosity as to why most if not all of the results can be obtained without seemingly the need to resort to Hilbert space. This goes against the prevailing orthodoxy that appears to insist that quantum mechanics cannot be done except in the context of a Hilbert space. Yet there have been other voices raised against the necessity of Hilbert space. Von Neumann himself wrote to Birkoff (1966) writing "I would like to make a confession which may seem immoral: I do not believe absolutely in Hilbert space any more." (A detailed discussion of why von Neumann made this comment can be found in Rédei 1996).

But there are more important reasons why an algebraic approach has advantages. As Dirac (1965) has stressed, when algebraic methods are used for systems with an infinite number of degrees of freedom (e.g., field theory), we can obtain solutions to some physical problems that give no solution in the usual Schrödinger picture. This possibility arises because in these systems the equivalence between the Schrödinger representation and the Heisenberg representation no longer holds. Indeed in field theory there is no longer one unique representation. There exist many inequivalent representations and these give rise to more general structures as has been pointed out by Emch (1972) and more recently in Haag (1992).

These works are of a more mathematical nature but from the point of view of physics what emerges is that the density operator, ρ, rather than the wave function, plays a fundamental role. This means we have a natural way of describing more general quantum processes that involve, for example, thermal systems. Systems described by wave functions then correspond to the special case when $\rho^2 = \rho$ which means the system is effectively at zero temperature. But it is not merely a question of the appearance of non-zero temperatures. As Prigogine (1994) has already pointed out, nonintegrable systems give rise to diffusive terms that correspond to the symmetry breaking in time. Such systems cannot be described in terms of wave functions. The only possible way of describing such systems is through the density operator.

---

[1] These algebras are called generalised Clifford algebras by Morris (1967)

While all the important new features that arise in the algebraic approach involve an infinite number of degrees of freedom even in the case of systems with a finite number of degrees of freedom we find new features that are rarely discussed in the physics literature. One such example, which we shall discuss later, is the appearance of algebraic spinors. We show that these spinors are a generalisation of the ordinary spinor, which have new mathematical implications that are ignored in the usual approach.

This generalisation is not only an enrichment of the mathematics. Following on from the work of Frescura and Hiley (1980a, 1980b), Monk and Hiley (1998) have pointed out that the algebra provides a very different perspective on the nature of quantum processes. In fact by emphasising the purely algebraic aspect of the approach we remove the distinction between operator and operand and this allows us to provide an interpretation in terms of the notion of process rather than in terms of particles and/or fields in i teraction. These ideas were developed to explore connections with non-commutative geometry but we will not develop these ideas here (See Hiley 1991).

Indeed it is not the purpose of this paper to attempt to justify this more general and speculative position here. Rather we want to explore the mathematical consequences of adopting this more general position by bringing out its relationship to the usual approach, a feature that is generally lost in the many details provided, for example, in excellent two volume treatise by Bratteli and Robinson (1979). In particular we want to present a simpler approach through which we can see what results depend only on the structure of the abstract algebra, what features depend on a specific representation, usually a matrix representation, and what features require the construction of a Hilbert space.

To maintain generality we start by recalling the main properties of an abstract * algebra. We then show a how a measure, ω, called a weight is introduced, which plays the role of the state. ω is a functional, mapping elements of the algebra onto the real field. It is not difficult to show that this is equivalent to introducing the density operator in the usual approach

To motivate the approach we first explore the orthogonal Clifford algebra by showing how primitive idempotents play a crucial role in this approach. We use these idempotents can be used to construct minimal left and right ideals. These ideals are spanned by algebraic spinors and these elements play the role of 'wave functions' in the Hilbert space formalism. But it must be emphasised these elements are contained entirely within the algebra and no external vector space is needed. We have already stressed this point in Frescura and Hiley (1980a). In this paper we show how these same ideals are generated in the standard algebraic approaches described in Haag (1992), and in Emch

(1972). As the aim of this paper is mainly pedagogic, we rely heavily on the theorems proved in those volumes, particularly the latter and will not reproduce them here. We then go on to illustrate how these methods can be to specific example that use two of the simplest Clifford algebra $R_{1,1}$ and H. $R_{1,1}$ is the Clifford algebra of a relativistic space-time with one space dimension while H is the quaternion algebra. We choose these not for physical reasons but because they are the simplest examples in which to demonstrate the principles involved. We also show how the density operator is constructed and the role it plays in the whole structure.

We then widen the discussion to the discrete Weyl algebra C^n (Weyl 1931 and Monk and Hiley 1993) ultimately generalising to the symplectic Clifford algebra (Crumeyrolle 1990). This enables us to extend the discussion to include the Heisenberg algebra.

We bring out clearly what is involved in this structure by first exploring C^3. In the mathematics literature this gives the structure of the nonions first introduced as a generalisation of the quaternion many years ago by Sylvester (1884). As the generalisation to higher n is straight forward but tedious we finally discuss the limit as $n \rightarrow \infty$ to include the Heisenberg algebra to make contact with standard quantum mechanics. In doing this we show how the algebraic approach is related to the Hilbert space approach.

Finally we show how the density matrix can be written as a vector in a higher dimensional space, which can be mapped into a Hilbert space. This is the so-called GNS construction, which arises in a very straightforward manner in the algebraic approach. We are then immediately able to connect up with the thermal field theory approach of Umezawa (1993) that exploits the rich properties of bialgebras. Physically this approach allows a description of the thermal properties of quantum processes.

2 The Algebra.

Since our main intention of the paper is pedagogical we begin by outlining the properties of the general algebra within which we will work. Recall that a linear algebra, $\mathtt{A}$, is a vector space with (i) addition $A + B \in \mathtt{A}$, (ii) multiplication by a scalar, $\lambda A \in \mathtt{A}$, λ is an element of the real or complex field, (iii) together with a product $AB \in \mathtt{A}$. We will assume the product to be associative. Furthermore we will initially consider the algebra to be of finite rank, n, with a finite basis $\{e_i\}$.

Since we require the analogue of a Hermitian conjugate we assume the algebra, A, is equipped with an involutive anti-automorphism defined by *: A → A with the following properties :

For $\forall A, B \in \mathtt{A}$

$$(A^*)^* = A, \qquad (\lambda A)^* = \lambda^* A^* \quad \lambda \in C$$

$$(A + B)^* = A^* + B^* \qquad (AB)^* = B^* A^*$$

We will further assume the following two requirements are satisfied,

(1) For each $A \in \mathtt{A}$, $\exists\, B$ in A such that $A^*A = B$

(2) $A^*A = 0 \Rightarrow A = 0.$

So as not worry too much about the abstract nature of this algebra, it is useful to keep in mind a matrix algebra with Hermitian conjugation defined.

3. The Density Operator.

3.1 The Algebraic preliminaries.

We need some way of obtaining numbers from this algebra that will allow us to identify with expectation values used in quantum mechanics. We use a analogous method to that used in set theory by introducing a functional ω such that

$$\omega: A \to \Re \text{ or } C \qquad \forall A \in \mathtt{A} \qquad \text{such that} \qquad \omega(A) = \alpha, \qquad \alpha \in \Re \text{ or } C$$

ω is a positive linear functional (called the expectation or state functional) satisfying,

$$\omega(\alpha A + \beta B) = \alpha\omega(A) + \beta\omega(B); \qquad \omega(\lambda A) = \lambda\omega(A) \qquad \text{and} \qquad \omega(A^*A) \geq 0$$

If the algebra has a unit, I, then we choose $\omega(I) = 1$. This means that the state is normalsed. The collection of all linear functionals over A forms the dual space A* of the algebra.

Now for $\forall A \in \mathtt{A}$, we have a real or complex number $\omega(A)$ depending linearly on A with $\omega(A^*A) \geq 0$. This enables us to define a Hermitian scalar product through

$$\langle A|B\rangle = \omega(A^*B) \qquad A, B \in \mathtt{A}. \qquad (1)$$

The collection of all positive linear functionals over A is the positive cone $\mathtt{A}^{*(+)}$ of the dual space A* of the algebra.

In A* ω is called 'extremal' if it cannot be decomposed into a linear combination of two others. These extremal functionals are called 'pure states'. Thus in general we can form

$$\omega(A) = \sum_{i=1}^{r} \lambda_i \omega_i(A) \qquad \text{with} \qquad \sum_i \lambda_i = 1$$

where ω_i are a set of extremal states.

As the algebras we consider in this paper are non-nilpotent, they have a set of idempotents. We can relate the extremal states to the primitive idempotents of the algebra in the following way. If $\{\varepsilon_i\}$ are a set of primitive idempotents[2], we can write $1 = \sum_{i=1}^{r} \varepsilon_i$ where $\varepsilon_i\varepsilon_j = \delta_{ij}\varepsilon_j$. Since $\omega(I) = 1$ we have

$$\omega(1) = \sum_{i=1}^{r} \lambda_i \omega_i(1) = \sum_{i=1}^{r} \lambda_i \omega_i\left(\sum_{j=1}^{r} \varepsilon_j\right) = \sum_{i,j=1}^{r} \lambda_i \omega_i(\varepsilon_j)$$

This defines the condition $\omega_i(\varepsilon_j) = \delta_{ij}$ so that the condition $\sum_i \lambda_i = 1$ is satisfied. In this way we establish a one-to-one correspondence between extremal states and primitive idempotents in the algebra.

We emphasise that through out this discussion we remain entirely within the algebra A even though the precise choice of the functional ω is left open at this stage. However if we are thinking in terms of a matrix algebra, then we can think of the linear functional as defining the trace. However it is not necessary to find a matrix representation of an algebra to define the invariant concept of a trace. For a finite algebra the trace can be evaluated by examining the coefficients of the characteristic polynomial of the algebra. A detailed discussion of this technique as applied to the algebras treated in this paper will be found in Frescura and Hiley (1981).

---

[2] An idempotent ε is primitive iff there does not exist two idempotents ε_1, ε_2 such that $\varepsilon = \varepsilon_1 + \varepsilon_2$. An idempotent is primitive iff $\varepsilon A \varepsilon = \lambda\varepsilon$ where $A \in \mathtt{A}$ and $\lambda \in Z$ the centre of A.

3.2 The Physics.

In quantum mechanics we are interested in expectation values. If the system is described by a pure state with wave function $\psi(r, t)$ then the expectation value is defined through the inner product

$$\langle A \rangle = \langle \psi | A | \psi \rangle = \int \psi^*(r,t) A \psi(r,t) d^3 r$$

where the assumption of the continuity of the wave function $\psi(r,t)$ forces us into an abstract Hilbert space. In what follows we will not make this assumption and will work completely in the algebra. Thus Hilbert space in not essential to deal with spin systems such as the Pauli spinor, the Dirac spinor and the twistor.

In the usual approach a system in a mixed state requires a density matrix. In this case the expectation value $\langle A \rangle = \sum_{i=1}^{r} \lambda_i \langle \psi_i | A | \psi_i \rangle$ can be written as

$$\langle A \rangle = Tr(\rho A) \qquad (2)$$

where ρ is the density operator with the properties that $\rho^\dagger = \rho$, $Tr(\rho A^* A) \geq 0$ and $Tr(\rho) = 1$.

Now in the usual approach using Hilbert space we always have to make a separation between the algebra of observables and the vector space on which the observables operate, but when the system is described by a finite algebra A, there is no need to introduce this extra external space. Everything can be done within the algebra as there is a suitable vector space already contained completely within the algebra. Indeed there are a set of subspace of this vector space and these correspond to the minimal left ideals of the algebra. There is also a corresponding dual vector subspace, namely, the minimal right ideals. It is through these ideals that one can describe the physical properties of the physical system without leaving the algebra. (See Frescura and Hiley 1980a, 1980b, Benn and Tucker 1983 and Monk and Hiley 1998)

Let us recall how this can be done. Recall that a left ideal, I_L consists of a set of elements K such that

$$I_L = \{K \in \mathrm{A} \mid AK \in I_L \mid \forall\, A \in \mathrm{A}\} \qquad (3)$$

These are just the algebraic spinors discussed in Frescura and Hiley (1980a). They are generated by the primitive idempotents introduced above. For example, if ε_i is a primitive idempotent, then a minimal left ideal I_L is generated as follows;

$$K = A\varepsilon_i \qquad \forall A \in \mathrm{A} \text{ and } K \in I_L$$

Now a right ideal I_R is defined through

$$I_R = \{K \in \mathrm{A} \,|\, KA \in I_R \,|\, \forall A \in \mathrm{A}\} \tag{4}$$

So that a right ideal I_R can also be generated by ε_i giving

$$K' = \varepsilon_i A \qquad \forall A \in \mathrm{A} \text{ and } K' \in I_R$$

Since $(\varepsilon_i)^* = \varepsilon_i$ we can generate a right ideal that is dual to the left ideal and is defined through

$$K^* = \varepsilon_i A^* \qquad \forall A^* \in \mathrm{A}$$

Thus if $A \in I_L$ and $B^* \in I_R$ then equation (1) becomes

$$\omega(B^* A) = \langle B|A\rangle \tag{5}$$

where $\langle B|A\rangle \in \mathrm{C}$.

In order to make contact with the expectation values used in physics, the pure state 'wave function' is thus replaced by an element of the minimal left ideal and its Hermitian conjugate is replaced by an appropriate element of the right ideal. Then if $B_\psi \in I_L$ and $B_\psi{}^* \in I_R$,

$$\langle A\rangle = \omega_\psi\left(B^*{}_\psi A B_\psi\right) \equiv \langle\psi|A|\psi\rangle = Tr\left(\rho_\psi A\right) \tag{6}$$

where ρ_ψ is the density operator for the pure state. We can now construct a density operator as a element in the algebra as follows:

$$\langle A\rangle = \omega_\psi\left(B^*{}_\psi A B_\psi\right) = \omega_\psi\left(A B_\psi B^*{}_\psi\right) = \omega_\psi\left(A\rho_\psi\right) = \omega_\psi\left(\rho_\psi A\right)$$

Here the density operator is

$$\rho_\psi = B_\psi B_\psi^* \tag{7}$$

so that the density operator can be written as an element of the algebra, A. We can also regard B as the 'square root' of the density operator, $B = \sqrt{\rho}$.

Let us make explicit the role of the primitive idempotent by writing

$$\rho_i = A\varepsilon_i A^* \qquad \forall A \in \mathrm{A}$$

Thus we see that the density operator is an element of a two-sided ideal subject to the conditions that ρ_i is a positive and $\omega(\rho_i) = 1$. This puts a normalisation constraint on A.

We can check that this density operator actually does correspond to a pure state by showing directly that ρ_i itself is idempotent. In fact

$$\rho_i^2 = (A\varepsilon_i A^*)(A\varepsilon_i A^*) = A(\varepsilon_i A^* A\varepsilon_i)A = \lambda A\varepsilon_i A^* = \lambda\rho_i$$

where $\lambda \in \Re$ or C. ρ_i can always be re-defined to absorb λ.

In the case of the mixed state we have

$$\langle A\rangle = \sum_{i=1}^{r} \lambda_i \omega_i (B_i^* A B_i) = \sum_{i=1}^{r} \lambda_i Tr(\rho_i A) \equiv Tr(\rho A) \tag{8}$$

where the density operator takes the form

$$\rho = \sum_{i=1}^{r} \lambda_i B_i B_i^* = \sum_{i=1}^{r} \left(\sqrt{\lambda_i} B_i\right)\left(\sqrt{\lambda_i} B_i^*\right) \tag{9}$$

If we are able to write the density operator in the form $\rho = DD^*$ and we will show how this can be done later, then we can write

$$\langle A\rangle = \omega\left(ADD^*\right) = \omega\left(D^* AD\right) \equiv \langle\Psi|A|\Psi\rangle \tag{10}$$

Thus even in a mixed state we can describe the system by a single vector D. This is equivalent to replacing the density matrix as a single 'wave function' $|\Psi\rangle$ in the usual Hilbert space formalism. Thus we have the possibility of describing a thermal system by a wave function implying that this wave function is a function of temperature. In other

words the wave function becomes a function of temperature. This is what is called the GNS construction (see for example Emch 1972).

From these results we see that the algebraic approach gives primary significance to the density operator and that this operator lies entirely within the algebraic structure itself. The 'wave function' is replaced by an element of the left ideal, which, as we have seen, is an element of the algebra. Thus at least for finite algebras the Hilbert space is an external auxiliary feature that is not essential to describe quantum systems. Thus we can regard the wave function simply as a device to enabling us to calculate expectation values using familiar mathematics and not as some feature describing the state of the system.

This whole approach allows us to start from the more general mixed state, seeing the pure state as a particular rather simple state. Methodologically this is much more satisfactory than the usual approach because we are not faced with the need to generalise the formalism when we find we have to deal with more general systems that require 'outside' factors to somehow destroy coherence and form mixed states. Indeed the physicist's emphasis on the wave function as being a description of the state of a quantum system and therefore a primitive element of the description may be too restrictive (See Prigogine 1994).

A more general approach based on the algebra as suggested by Monk and Hiley (1998) has the possibility of avoiding some of the interpretation problems in quantum mechanics. However this comment should not be taken to imply that this new approach will enable us to solve all the problems of interpretation. It does not. What it does is to open up another way of looking at quantum phenomena in which the algebra itself is playing a primary role. In other words singling out the density operator as the primary feature of the mathematics opens up the possibility of giving a different meaning to the formalism (See Fernandes and Hiley 1997 and Brown and Hiley 2000).

As I have argued elsewhere, this approach suggests that a notion of process rather than particles/fields-in-interaction should be taken as basic (Hiley 1995). In a process based philosophy there are no separately existing objects. They are invariants of the total unfolding process. Here the wave function is regarded as a transition probability amplitude rather than a function of state of some system. As Bohr insisted, there is no sharp separation between subject and object so we should have no sharp separation between operator and operand. The mathematics of the algebraic approach reflects these notions and makes no such distinction. Thus we do not have 'operators' playing the role of 'observables' which are to 'act' on vectors in a different abstract space. All aspects of the process are united in the same structure described by elements of the same algebraic

structure, elements that only have meaning in the context of the total algebra. This approach has some interesting consequences that have already been discussed in some detail by Hiley and Monk (1993) and will not be discussed further here. The emphasis in this paper will be on those aspects of the algebraic structure that are relevant to this general discussion.

4. Minimal Left Ideals, Algebraic Spinors.

4.1 Pure States.

Let us now illustrate how these ideas work by considering a finite, linear *algebra on which a state ω is defined. We require to find a set of minimal left ideals. We will not follow the techniques based on the characteristic polynomial which have already been outlined in Frescura and Hiley (1978). Rather we follow the method that used by Emch (1972) and is essentially based on the fact that $\omega_i(\varepsilon_j) = \delta_{ij}$ as discussed earlier. They produce exactly the same results but this method is the one conventionally used in algebraic quantum mechanics. (See, for example, Haag 1992)

We can construct left ideals by solving the equation $\omega(B^*K) = 0$ for K and for all $B^* \in$ A. Here we are using the definition

$$I_{L\omega} = \left\{ K \in \; \middle| \omega\left(B^* K\right) = 0, \forall B \in \; \right\} \tag{11}$$

to construct the left ideals. To show this definition does in fact produce a left ideal take any A and B in A and any $K \in I_{L\omega}$ and use associativity to establish

$$\omega\left(A^*(BK)\right) = \omega\left(\left(B^* A\right)^* K\right) = 0 \tag{12}$$

showing that $I_{L\omega}$ is a left ideal in A. The left ideal generated in this way is sometimes called the Gel'fand ideal.

Similarly a right ideal can be generated by the sub-set

$$I_{R\omega} = \left\{ R \in \; \middle| \omega\left(RB^*\right) = 0, \forall B \in \; \right\}$$

4.2. A Simple Example to illustrate these Ideas.

As the main purpose of this paper is pedagogic, I will now demonstrate how these ideas work in a simple real Clifford algebra $R_{1,1}$. This example is not without physical meaning as it forms the basis of relativity in one spatial dimension, and one time dimension. It is the background structure to spin networks and relativity discussed by Kauffman (1991).

The algebra is characterised by the multiplication table below

	1	e_1	e_2	e_{12}
e_1	1	e_{12}	e_2
e_2	$-e_{12}$	-1	e_1
e_{12}	$-e_2$	$-e_1$	1

with $e_1^* = e_1 \; ; \; e_2^* = -e_2 \; ; \; e_{12}^* = e_{12}.$

Define the state ω_1 by

$$\omega_1(\alpha 1 + \beta e_1 + \gamma e_2 + \delta e_{12}) = \alpha - \delta \tag{13}$$

It is easy to show using $\omega_1(B^*A) = 0$, that

$$\omega_1\left(\left(\alpha 1 + \beta e_1^* + \gamma e_2^* + \delta e_{12}^*\right)\left(a(1 + e_{12}) + b(e_1 + e_2)\right)\right) = 0$$

confirming that ω_1 produces an element in a left ideal Ψ_{L1} defined by

$$\Psi_{L1} = a\varepsilon_{11} + b\varepsilon_{21} \tag{14}$$

where $\varepsilon_{11} = \frac{1}{2}(1 + e_{12})$ and $\varepsilon_{21} = \frac{1}{2}(e_1 + e_2)$.

The element in the corresponding right ideal can be shown to be

$$\Psi_{R_1} = a^*\varepsilon_{11} + b^*\varepsilon_{12} \tag{15}$$

where $\varepsilon_{11} = \frac{1}{2}(1 + e_{12})$ and $\varepsilon_{12} = \frac{1}{2}(e_1 - e_2)$

Then we can form

$$\omega_1(\Psi_{R1}{}^*\Psi_{L1}) = \langle \Psi_{R1} | \Psi_{L1} \rangle = |a|^2 + |b|^2$$

which is unity for a normalised state.

We now construct a density operator from elements in these ideals. Recalling the definition of the density operator defined in section 4.0, we find that this density operator is

$$\rho_1 = \Psi_{L1}\Psi_{R1} = \tfrac{1}{2}\left[\left(|a|^2 + |b|^2\right)1 + \left(ab^* + a^*b\right)e_1 + \left(a^*b - ab^*\right)e_2 + \left(|a|^2 - |b|^2\right)e_{12}\right]$$

$$= |a|^2\varepsilon_{11} + ab^*\varepsilon_{12} + a^*b\varepsilon_{21} + |b|^2\varepsilon_{22} \qquad (16)$$

We can generate a second set if elements in second left ideal by using the state

$$\omega_2(\alpha 1 + \beta e_1 + \gamma e_2 + \delta e_{12}) = \alpha + \delta \qquad (17)$$

so that

$$\omega_2\left(\left(\alpha 1 + \beta e_1^* + \gamma e_2^* + \delta e_{12}^*\right)\left(c(1 - e_{12}) + d(e_1 - e_2)\right)\right) = 0$$

producing the element of the left ideal, Ψ_{L2} defined by

$$\Psi_{L2} = c\varepsilon_{22} + d\varepsilon_{12}. \qquad (18)$$

where $\varepsilon_{22} = \frac{1}{2}(1 - e_{12})$ and $\varepsilon_{12} = \frac{1}{2}(e_1 - e_2)$.

The corresponding element in the right ideal is

$$\Psi_{R2} = c^*\varepsilon_{22} + d^*\varepsilon_{21} \qquad (19)$$

where $\varepsilon_{22} = \frac{1}{2}(1 - e_{12})$ and $\varepsilon_{21} = \frac{1}{2}(e_1 + e_2)$.

Then we find

$$\omega_2(\Psi_{R2}{}^*\Psi_{L2}) = \langle \Psi_{R2} | \Psi_{L2} \rangle = |c|^2 + |d|^2$$

while the density operator is

$$\rho_2 = \Psi_{L2}\Psi_{R2} = \tfrac{1}{2}\left[\left(|c|^2 + |d|^2\right)1 + \left(cd^* + c^*d\right)e_1 + \left(cd^* - c^*d\right)e_2 - \left(|c|^2 - |d|^2\right)e_{12}\right]$$

$$= |d|^2\varepsilon_{11} + c^*d\varepsilon_{12} + cd^*\varepsilon_{21} + |c|^2\varepsilon_{22} \qquad (20)$$

We see immediately that the two left ideals are generated by the idempotents

$$\varepsilon_{11} = \tfrac{1}{2}(1+e_{12}) \qquad \text{and} \qquad \varepsilon_{22} = \tfrac{1}{2}(1-e_{12})$$

so that $\varepsilon_{11} + \varepsilon_{22} = 1$ and $\varepsilon_{11}.\,\varepsilon_{22} = 0.$

Since these idempotents are primitive we have generated two minimal left ideals. These are the results presented in Frescura and Hiley (1980a). A more lengthy discussion of the Pauli Clifford will be found in Frescura and Hiley (1978).

4.3 A Matrix Representation.

These results can be made clearer if we consider the following matrix representation although these matrix representations are not necessary in general;

$$\pi(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \qquad \pi(e_2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \qquad \pi(e_{12}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad (21)$$

In terms of matrices, the elements of the minimal left ideals become

$$\pi(\Psi_{L1}) = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \qquad \text{and} \qquad \pi(\Psi_{L2}) = \begin{pmatrix} 0 & d \\ 0 & c \end{pmatrix} \qquad (22)$$

The appearance of two minimal left ideals may at first sight seems surprising since in physics there is only one column vector for spin[3]. In section 8 we will show exactly how these two spinors are merged into one. At this stage we merely remark that the two spinors are equivalent under the inner automorphsim defined by

[3] It should be noted that minimal left ideals are not always represented as single column matrices. For example $\pi_2(\Psi_{L1}) = \tfrac{1}{2}\begin{pmatrix} a-b & a-b \\ a+b & a+b \end{pmatrix}$ is a minimal left ideal.

$$\pi(e_1)=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

The corresponding elements of the minimal right ideals become

$$\pi(\Psi_{R_1})=\begin{pmatrix}a^* & b^*\\ 0 & 0\end{pmatrix} \quad \text{and} \quad \pi(\Psi_{R_2})=\begin{pmatrix}0 & 0\\ d^* & c^*\end{pmatrix} \tag{23}$$

Finally let us return to the definition of the state ω_1 and see how it fits in with the matrix representation. In this case the density operator takes the special form ρ_1 shown above so that

$$\omega_1(\alpha 1+\beta e_1+\gamma e_2+\delta e_{12})=Tr\{\rho_1(\alpha 1+\beta e_1+\gamma e_2+\delta e_{12})\}=\alpha-\delta \tag{24}$$

whereas for ω_2 the density operator takes the form ρ_2 so that

$$\omega_2(\alpha 1+\beta e_1+\gamma e_2+\delta e_{12})=Tr\{\rho_2(\alpha 1+\beta e_1+\gamma e_2+\delta e_{12})\}=\alpha+\delta \tag{25}$$

Thus we find consistency since equations (24) and (25) are identical to equations (13) and (17). This means that the state defined by ω_1 generates the left ideal generated by ε_2, and vice versa. The reason for this is as follows. Since $\omega_1(B^*K) = 0$ for $\forall\ B \in \mathrm{A}$, we must have $\omega_1(K^*K) = 0$ for $\forall\ K \in \mathrm{I}_{L1}$. Since $\omega_1(\varepsilon_1{}^*\varepsilon_1) \neq 0$, $\varepsilon_1 \notin \mathrm{I}_{L1}$. Thus the left ideal generated by $\omega_i(B^*K) = 0$ is always the complement of the ideal generated by ε_i. We will discuss this point again in section 8.1.

4.4 Mean Values.

As we have seen above we can write

$$\langle A\rangle=\omega_1(\Psi_{R1}A\Psi_{L1})$$

Now if we write

$$A=a_0 1+a_1e_1+a_2e_2+a_{12}e_{12}$$

Then using the above expressions for Ψ_{L1} and Ψ_{R1} above we find

$$(\Psi_{R1}A\Psi_{L1})=\left[\left(|a|^2+|b|^2\right)a_0+\left(a^*b+b^*a\right)a_1+\left(b^*a-a^*b\right)a_2+\left(|a|^2-|b|^2\right)a_{12}\right]\tfrac{1}{2}(1+e_{12})$$

while

$$\langle A\rangle = \omega_1(\Psi_{R1}A\Psi_{L1}) = \left[\left(|a|^2+|b|^2\right)a_0 + \left(a^*b+b^*a\right)a_1 + \left(b^*a-a^*b\right)a_2 + \left(|a|^2-|b|^2\right)a_{12}\right] \quad (26)$$

We can also write

$$\langle A\rangle = \omega_1(\Psi_{R1}A\Psi_{L1}) = \omega_1(A\Psi_{L1}\Psi_{R1}) = \omega_1(A\rho_1) = Tr(\rho_1 A)$$

Indeed one can evaluate $\omega_1(A\rho_1)$ and show it is equal to the result calculated in the last equation.

5. Mixed States and the GNS Construction.

In order to illustrate how to deal with mixed states for the algebra $R_{1,1}$, we start with the matrix representation defined in section 4.3 and choose the specific density matrix

$$\pi(\rho) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

It can quickly be checked that ρ in fact corresponds to mixed state by showing $\rho^2 \neq \rho$. Remembering that ω is equivalent to taking the trace, it is straight forward to show that

$$\omega_\rho(A) = \omega_\rho(a_0 1 + a_1 e_1 + a_2 e_2 + a_{12}e_{12}) = (\lambda_1+\lambda_2)a_0 + (\lambda_1-\lambda_2)a_{12} \quad (27)$$

Let us now try to find the Gel'fand ideal using the technique outlined in section 4.1 and 4.2. We start by writing this ideal in the general form

$$\Psi_L = z + z_1 e_1 + z_2 e_2 + z_{12}e_{12}$$

Then we need to solve the equation

$$\omega_\rho\left((a_0 + a_1 e_1 + a_2 e_2 + a_{12}e_{12})(z + z_1 e_1 + z_2 e_2 + z_{12}e_{12})\right) = 0$$

with

$$\omega_\rho(a_0 1) = a_0(\lambda_1+\lambda_2);\quad \omega_\rho(be_1) = 0;\quad \omega_\rho(ce_2) = 0;\quad \omega_\rho(de_{12}) = d(\lambda_1-\lambda_2)$$

We can then show that the equation we have to solve is

$$a_0\left[(\lambda_1+\lambda_2)z+(\lambda_1-\lambda_2)z_{12}\right]+a_1\left[(\lambda_1+\lambda_2)z_1+(\lambda_1-\lambda_2)z_2\right]+$$
$$a_2\left[(\lambda_1+\lambda_2)z_2+(\lambda_1-\lambda_2)z_1\right]+a_{12}\left[(\lambda_1+\lambda_2)z_{12}+(\lambda_1-\lambda_2)z\right]=0$$

This equation has no non-trivial solution except for $\lambda_1 = 1$, $\lambda_2 = 0$; and $\lambda_1 = 0$, $\lambda_2 = 1$ which are the results we obtained in section 4.2. Hence we cannot extend these methods to general values of λ and therefore we have to use an alternative method, which we will discuss below. However before going on to explore a more general method let us first explore how the mixed state can be handles within $R_{1,1}$.

In section 3.2 equation (9), we showed that the density matrix could be written in the form

$$\rho=\sum_{i=1}^{r}\sqrt{\lambda_i}B_i\sqrt{\lambda_i}B_i^{\dagger}$$

which in this example can be written as

$$\rho=\rho_1+\rho_2=\Psi_{R1}\Psi_{L1}+\Psi_{R2}\Psi_{L2}$$

where

$$\Psi_{L1}=\sqrt{\lambda_1}\tfrac{1}{2}(1+e_{12})=\sqrt{\lambda_i}\varepsilon_{11} \qquad \text{with} \qquad \Psi_{R1}=\sqrt{\lambda_1}\tfrac{1}{2}(1+e_{12})=\sqrt{\lambda_i}\varepsilon_{11}$$

and

$$\Psi_{L2}=\sqrt{\lambda_2}\tfrac{1}{2}(1-e_{12})=\sqrt{\lambda_2}\varepsilon_{22} \qquad \text{with} \qquad \Psi_{R2}=\sqrt{\lambda_2}\tfrac{1}{2}(1-e_{12})=\sqrt{\lambda_2}\varepsilon_{22}$$

so that

$$\rho=\rho_1+\rho_2=\lambda_1\varepsilon_{11}+\lambda_2\varepsilon_{22} \tag{28}$$

Then

$$\langle A\rangle=Tr(\rho A)=\omega_1(\rho_1 A)+\omega_2(\rho_2 A)$$

gives

$$\langle A\rangle=(\lambda_1+\lambda_2)a_0+(\lambda_1-\lambda_2)a_{12} \tag{29}$$

This then agrees with the result calculated using the matrix representation above. Thus we see that we use both minimal ideals in constructing the density operator. This result has already been used for the Dirac Clifford algebra by Frescura and Hiley (1987)

5.1 GNS Construction.

As we indicated in section 3.2 we would like to construct a D so that we can write the density operator as $\rho = DD^{\dagger}$ which in turn allows us to write

$$\langle A\rangle = \omega\left(D^{\dagger}AD\right) = \omega\left(DD^{\dagger}A\right) = \omega(\rho A) = \langle\Psi|A|\Psi\rangle$$

so that D is the algebraic equivalent of $|\Psi\rangle$. This would enables to find the algebraic background to the GNS construction.

From equation (28) together with the fact that ε_{11}, ε_{22} are the orthogonal primitive idempotents satisfying $\varepsilon_{ii}\varepsilon_{jj} = \delta_{ij}\varepsilon_{jj}$, we can immediately write

$$D = \sum_{i=1}^{2}\sqrt{\lambda_i}\,\varepsilon_{ii} \tag{30}$$

so that

$$\langle A\rangle = \omega\left(\sum_{r=1}^{2}\sqrt{\lambda_i}\,\varepsilon_{ii}A\sum_{r=1}^{2}\sqrt{\lambda_i}\,\varepsilon_{ii}\right) = (\lambda_1+\lambda_2)a_0 + (\lambda_1-\lambda_2)a_{12}$$

which is the result produced in equation (29). Thus we have obtained exactly the same result as that obtained by matrix representation used in section 4.2. It is now obvious how to generalise this. We simply write

$$D = \sum_{i=1}^{r}\sqrt{\lambda_i}\,\varepsilon_{ii} \tag{31}$$

which immediately gives

$$\langle A\rangle = \omega\left(D^{\dagger}AD\right) = \omega(\rho A) \equiv \langle\Psi|A|\Psi\rangle$$

5.2 The Matrix Representation

The GNS construction is usually presented as a matrix representation (See Emch 1972). We can immediately construct a matrix representation if we first notice that any matrix transformation of the form $C\rho D$ can be written in the form $(C\otimes D)\hat{\rho}$ where $\hat{\rho}$ is the $n\times n$ density matrix written as a column with n^2 entries. But we are interested in forming

$$\langle A\rangle = Tr(\rho A) = Tr\left(DD^{*}A\right) = Tr\left(D^{*}AD\right) = Tr\left(\pi\left(D^{*}\right)\pi(A\otimes 1)\pi(D)\right) \quad (32)$$

where $\pi(D)$ is a column with n^2 entries. This can be written as

$$\pi(D) = \sum_{j=1}^{r}\sqrt{\lambda_j}\pi(\varepsilon)_{jj} = \sum_{j=1}^{r}\left(B_j\right) \quad (33)$$

In the matrix representation we can write

$$\pi\left(B_j\right) = \begin{pmatrix} 0 \\ . \\ 0 \\ \sqrt{\lambda_j} \\ 0 \\ . \\ 0 \end{pmatrix}$$

That is where $\pi\left(B_j\right)$ is the $1\times n^2$ column matrix with $\sqrt{\lambda_j}$ appearing in the $s = [j(n+1)+1]$ element. Here λ_j is the j^{th} eigenvalue of the density matrix in the representation used in section 5. This enables us to introduce the column matrix $\pi(D)$ with $\sqrt{\lambda_j}\delta_{js}$ in the s^{th} element satisfying $s = [j(n+1)+1]$ as j runs from 1 to r. Then we can write

$$\langle A\rangle = \omega_{\rho}(A) = Tr(\rho A) = \omega\left(\pi(D^{*})(\pi(A)\otimes\pi(1))\pi(D)\right) \quad (34)$$

5.3 Specific example

Let us illustrate the above discussion by returning to the example in section 5.1. Here we find that $\pi(D)$ becomes

$$\pi(D) = \begin{pmatrix} \sqrt{\lambda_1} \\ 0 \\ 0 \\ \sqrt{\lambda_2} \end{pmatrix} \tag{35}$$

so that

$$\langle A \rangle = \langle \Psi | \hat{A} | \Psi \rangle = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & \sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} a_0 + a_{12} & a_1 - a_2 & & \\ a_1 + a_2 & a_0 - a_{12} & & \\ & & a_0 + a_{12} & a_1 - a_2 \\ & & a_1 + a_2 & a_0 - a_{12} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} \\ 0 \\ 0 \\ \sqrt{\lambda_2} \end{pmatrix}$$

$$\langle A \rangle = (\lambda_1 + \lambda_2) a_0 + (\lambda_1 - \lambda_2) a_{12}$$

which is identical to equation (27) showing that we can get exactly the same result as the standard method using the density matrix. The advantage of this method lies in the fact that when generalised to a Hilbert space construction, which we will discuss later, we have 'wave functions' depending upon temperature. This leads us to a more generalised quantum mechanics discussed by Umezawa (1993),

5.4 A more Physical Example.

To this point we have been considering mathematical structures per se. Let us now turn to see how these techniques in a physical problem. Consider a thermal spin system in an external magnetic field. A typical density operator can be written in the form

$$\rho = \frac{\exp[-\beta H]}{Tr \exp[-\beta H]} \tag{36}$$

where $\beta = 1/kT$ and H is the Hamiltonian $H = -B\sigma_3$. Since we are using Pauli spin matrices we must use the standard multiplication rules for Pauli spin matrices, i.e. the quaternion Clifford algebra, H and not $R_{1,1}$ as used above. Using the standard representation for the Pauli spin matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \text{and} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

we find

$$\rho = \begin{pmatrix} \frac{1+\tanh\gamma}{2} & 0 \\ 0 & \frac{1-\tanh\gamma}{2} \end{pmatrix} \tag{37}$$

then it is straightforward to show

$$\omega_\rho\left(a1 + b\sigma_x + c\sigma_y + d\sigma_z\right) = a + d\tanh\gamma \tag{38}$$

Let us now try to find the Gel'fand ideal using the technique outlined above. We start by writing this ideal in the general form

$$\Psi_L = z + z_1\sigma_1 + z_2\sigma_2 + z_3\sigma_3$$

Then we need to solve the equation

$$\omega\left(\left(a + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3\right)\left(z + z_1\sigma_1 + z_2\sigma_2 + z_3\sigma_3\right)\right) = 0$$

with

$$\omega(a1) = a; \quad \omega(b\sigma_1) = 0; \quad \omega(c\sigma_2) = 0; \quad \omega(d\sigma_3) = d\tanh\gamma$$

It is straightforward to show that the equation we eventually have to solve is

$$(a - a_3\tanh\gamma)z + (a_1 + a_2\tanh\gamma)z_1 + (a_2 - a_1\tanh\gamma)z_2 + (a\tanh\gamma + a_3)z_3 = 0.$$

Once again this equation has no non-trivial solution so we cannot used the methods outlined above in section 4 and we are forced to use the GNS construction. To do this we need a pair of primitive idempotents from the algebra H. These are

$$\varepsilon_{11} = \tfrac{1}{2}(1+\sigma_3) \qquad \text{and} \qquad \varepsilon_{22} = \tfrac{1}{2}(1-\sigma_3) \tag{39}$$

so that

$$D = \frac{\sqrt{\lambda_1}}{2}(1+\sigma_3) + \frac{\sqrt{\lambda_2}}{2}(1-\sigma_3) \tag{40}$$

If we write $A = a\mathrm{I} + b\sigma_1 + c\sigma_2 + d\sigma_3$ then

$$\langle A\rangle = \omega(A\rho) = \omega\left(D^{\dagger}AD\right) = a + d\tanh\gamma$$

which is the value found in equation (38). In the matrix representation discussed in 5.2 we find

$$\pi(D) = |\Psi\rangle = \begin{pmatrix} \sqrt{\dfrac{1+\tanh\gamma}{2}} & 0 \\ 0 & \sqrt{\dfrac{1-\tanh\gamma}{2}} \end{pmatrix} \quad \text{with} \quad \pi(\sigma) = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} \tag{41}$$

Again it is straight forward to check that

$$\omega_\rho(\pi(A)) = \langle\Psi|\pi(A)|\Psi\rangle = a + d\tanh\gamma$$

6. Umezawa's Approach.

I the above generalisation to mixed states we need to consider a product representation A ⊗ 1. It is of interest to compare the above approach with that introduced by Umezawa (1993).

Suppose we start with a general thermodynamic system with the partition function given by

$$Z(\beta) = Tr(\exp[-\beta H])$$

where $\beta = 1/kT$ and $H = H_0 - \mu N$. Here H_0 is the particle Hamiltonian and μ is the chemical potential. The ensemble average is given by

$$\langle A\rangle = \frac{Tr(\exp[-\beta H]A)}{Z(\beta)}$$

We want to find a state $|\Omega(\beta)\rangle$ such that

$$\langle A\rangle = \frac{Tr(\exp[-\beta H]A)}{Z(\beta)} = \langle\Omega(\beta)|\pi(A)|\Omega(\beta)\rangle$$

Suppose $H|n\rangle = E_n|n\rangle$, then

$$\langle A\rangle = Z^{-1}(\beta)\sum_n \langle n|A|n\rangle \exp[-\beta E_n] = \langle\Omega(\beta)|\pi(A)|\Omega(\beta)\rangle$$

Can we find $|\Omega(\beta)\rangle$? To this effect let us write $|\Omega(\beta)\rangle = \sum_n |n\rangle f_n(\beta)$ so that

$$\langle\Omega(\beta)|\pi(A)|\Omega(\beta)\rangle = \sum_{n,m} f_m^*(\beta) f_n(\beta)\langle m|A|n\rangle = Z^{-1}(\beta)\sum_n \langle n|A|n\rangle \exp[-\beta E_n]$$

which will be satisfied only if

$$f_m^*(\beta) f_n(\beta) = Z^{-1}(\beta)\exp[-\beta E_n]\delta_{mn}$$

However it is impossible to satisfy this relation if $f_n(\beta)$ are simply complex numbers. Umezawa suggests we write

$$f_n(\beta) = Z^{-1/2}(\beta)\exp[-\beta E_n/2]|n\rangle$$

Here $|n\rangle$ is a copy of the original set of kets. Then

$$f_m^*(\beta) f_n(\beta) = Z^{-1}(\beta)\exp[-\beta(E_n + E_m)/2]\delta_{mn}$$

This means we can write

$$|\Omega(\beta)\rangle = Z^{-1/2}\sum_n \exp[-\beta E_n/2]|n\rangle \otimes |n\rangle . \qquad (42)$$

We can show this is exactly the same vector as is obtained from the GNS construction. We will simply illustrate the similarity using the example of section 5.4.

It is straightforward to show that

$$Z^{-1/2}\exp[\beta E_1/2]|1\rangle \otimes |1\rangle = \sqrt{\frac{1+\tanh\gamma}{2}}\,|1\rangle \otimes |1\rangle$$

and

$$Z^{-1/2}\exp[\beta E_2/2]|2\rangle\otimes|2\rangle=\sqrt{\frac{1-\tanh\gamma}{2}}\,|2\rangle\otimes|2\rangle$$

Since

$$|1\rangle\otimes|1\rangle\rightarrow\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$$

and

$$|2\rangle\otimes|2\rangle\rightarrow\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

We find

$$|\Omega(\beta)\rangle=\begin{pmatrix}\sqrt{\frac{1+\tanh\gamma}{2}}\\0\\0\\\sqrt{\frac{1-\tanh\gamma}{2}}\end{pmatrix}$$

which is identical to equation (41), the result we obtained from the GNS construction.

What we see from this approach to the GNS construction is that we need to "double up" the algebra since we need two copies of the vector space. To put it more formally we need to form the bialgebra $\mathsf{A}\otimes\mathsf{A}$. Indeed this is exactly what Umezawa (1993) proposed in his approach to thermal field theory. Starting from the boson algebra generated by the set $\{1, a, a^\dagger\}$ where a is the boson destruction operator and $a^\dagger$ the corresponding creation operator, he introduced an additional pair of annihilation and creation operators $\tilde{a}$ and $\tilde{a}^\dagger$ which he attributes to a "ghost" field (Umezawa 1993a). Of course the idea of a ghost field is not very appealing from the physics point of view, however this approach opens up the possibility of using this doubling to account for quantum dissipation (Celeghini, Rasetti, and Vitiello 1992). Hiley and Fernandes (1997), following from Bohm and Hiley (1981), have shown this doubling of the number of degrees of freedom is also required in going from two point functions in configuration

space description to an algebraic phase space description. This gives an alternative view on dissipation. These ideas will be developed further in a forthcoming paper.

7. Extension to the Heisenberg Algebra.

All of the previous examples have used the orthogonal Clifford algebra in some form or other. What we would like to do now is to show how the whole procedure can be extended to the Heisenberg algebra. As pointed out in a previous paper (Hiley 2001), the main problem in following the above approach arises from the fact that the Heisenberg algebra is a nilpotent algebra of degree three under the product $[A, B]$. As a consequence of a well-known theorem, nilpotent algebras do not contain any non-trivial idempotents so it is not possible to construct any non-trivial left ideals.

We want to approach Heisenberg algebra in two steps. Rather than go to the full algebra immediately, we will start by considering the finite algebra introduced by Weyl (1930) which contains the Heisenberg algebra in the continuum limit. The Weyl algebra has an extremely simple structure. It is a finite polynomial algebra $C^2{}_n$ generated by the set of elements $\{I, e_1, e_2\}$ subject to

$$e_1e_2 = \omega\, e_2e_1 \qquad e_1{}^n = I, \qquad e_2{}^n = I, \qquad \text{with } \omega = \exp[2\pi i/n] \quad (43)$$

where n is an integer. This algebra has a long history being first explored by Sylvester in 1884 as an example of a generalisation of the quaternions. He called the elements nonions for $n = 3$ and n-ions more generally. What Weyl shows is that we can write $e_1 = \exp[i\xi P]$ and $e_2 = \exp[i\eta X]$ where $\xi = 2\pi/\delta pn$ and $\eta = 2\pi/\delta xn$ so that we begin to see the beginnings of the Heisenberg algebra. Indeed in the limit as $n \to \infty$ this algebra does approaches the Heisenberg algebra with X and P representing the position and momentum operators..

The importance of the algebra from our point of view is that it is not nilpotent. In order to find the idempotents using the method described in the section 4, equation (11) we need to generate the sub-set satisfying

$$_{\omega} = \left\{ K \in \quad \middle| \omega\left(B^{*}K\right) = 0, \forall B \in \quad \right\}$$

In order to see how this works in this example we again illustrate the procedure with an example.

7.1 The nonions.

We now take the simplest example of the structure described in the last sub-section, namely, $n = 3$, the nonions, and write down the explicit multiplication table explicitly.

1	e_1	e_2	e_{11}	e_{12}	e_{22}	e_{112}	e_{122}	e_{1122}
e_1	e_{11}	e_{12}	1	e_{112}	e_{122}	e_2	e_{1122}	e_{22}
e_2	$\omega^2 e_{12}$	e_{22}	ωe_{112}	$\omega^2 e_{122}$	1	ωe_{1122}	$\omega^2 e_1$	ωe_{11}
e_{11}	1	e_{112}	e_1	e_2	e_{1122}	e_{12}	e_{22}	e_{122}
e_{12}	$\omega^2 e_{112}$	e_{122}	ωe_2	$\omega^2 e_{1122}$	e_1	ωe_{22}	$\omega^2 e_{11}$	$\omega 1$
e_{22}	ωe_{122}	1	$\omega^2 e_{1122}$	ωe_1	e_2	$\omega^2 e_{11}$	ωe_{12}	$\omega^2 e_{112}$
e_{112}	$\omega^2 e_2$	e_{1122}	ωe_{12}	$\omega^2 e_{22}$	e_{11}	ωe_{122}	$\omega^2 1$	ωe_1
e_{122}	ωe_{1122}	e_1	$\omega^2 e_{22}$	ωe_{11}	e_{12}	$\omega^2 1$	ωe_{112}	$\omega^2 e_2$
e_{1122}	ωe_{22}	e_{11}	$\omega^2 e_{122}$	$\omega 1$	e_{112}	$\omega^2 e_1$	ωe_2	$\omega^2 e_{12}$

with $e_1{}^* = e_{11}$; $e_2{}^* = e_{22}$; $e_{12}{}^* = \omega^2 e_{1122}$; $e_{112}{}^* = \omega e_{112}$,

and $(e_i{}^*)^* = e_i$; $\omega^* = \omega^{-1}$.

We define the state by

$$\omega_1(a1) = a; \quad \omega_1(be_2) = b; \quad \omega_1(ce_{22}) = c; \quad \text{Rest} = 0.$$

Now we need to solve the equation $\omega_1(B^*K) = 0$, where

$$\begin{aligned} B^* &= \alpha 1 + \beta e_2{}^* + \gamma e_{22}{}^* + \delta e_1{}^* + \varepsilon e_{12}{}^* + \eta e_{122}{}^* + \theta e_{11}{}^* + \Delta e_{112}{}^* + \Omega e_{1122}{}^* \\ &= \alpha 1 + \beta e_{22} + \gamma e_2 + \delta e_{11} + \varepsilon \omega^2 e_{1122} + \eta \omega e_{112} + \theta e_1 + \Delta \omega e_{122} + \Omega \omega^2 e_{12} \end{aligned}$$

and

$$K = z1 + z_2 e_2 + z_{22} e_{22} + z_1 e_1 + z_{12} e_{12} + z_{122} e_{122} + z_{11} e_{11} + z_{112} e_{112} + z_{1122} e_{1122}.$$

It is tedious but straight forward to show that the solution of the equation $\omega_1(B^*K) = 0$ is

$$(\alpha + \beta + \delta)(z + z_2 + z_{22}) + (\delta + \varepsilon + \eta)(z_1 + z_{12} + z_{122}) + (\theta + \Delta + \Omega)(z_{11} + z_{112} + z_{1122}) = 0$$

which gives solution (A) as

$$z = 1; \qquad z_2 = \omega; \qquad z_{22} = \omega^2$$

$$z_1 = 1; \qquad z_{12} = \omega; \qquad z_{122} = \omega^2$$
$$z_{11} = 1; \qquad z_{112} = \omega; \qquad z_{1122} = \omega^2$$

so that the left ideal is

$$3\Psi_{L1}(A) = (1 + \omega e_2 + \omega^2 e_{22})a + (e_1 + \omega e_{12} + \omega^2 e_{122})b + (e_{11} + \omega e_{112} + \omega^2 e_{1122})c \qquad (44)$$

Solution (B) is

$$z = 1; \qquad z_2 = \omega^2; \qquad z_{22} = \omega$$
$$z_1 = 1; \qquad z_{12} = \omega^2; \qquad z_{122} = \omega$$
$$z_{11} = 1; \qquad z_{112} = \omega^2; \qquad z_{1122} = \omega$$

$$3\Psi_{L1}(B) = (1 + \omega^2 e_2 + \omega e_{22})a + (e_1 + \omega^2 e_{12} + \omega e_{122})b + (e_{11} + \omega^2 e_{112} + \omega e_{1122})c \qquad (45)$$

We see from these two solutions we have two idempotents

$\varepsilon_{11} = \frac{1}{3}(1 + \omega e_2 + \omega^2 e_{22})$ and $\varepsilon_{22} = \frac{1}{3}(1 + \omega^2 e_2 + \omega e_{22})$. We can find the third from

$$1 = \varepsilon_{00} + \varepsilon_{11} + \varepsilon_{22}$$

So that $\varepsilon_{00} = \frac{1}{3}(1 + e_2 + e_{22})$ from which we can generate the third left ideal by multiplying this idempotent from the left by each element of the algebra. For completeness we find this third ideal is explicitly

$$3\Psi_{L1}(C) = (1 + e_2 + e_{22})a + (e_1 + e_{12} + e_{122})b + (e_{11} + e_{112} + e_{1122})c \qquad (46)$$

It is now straight forward to use these three left ideals to construct density operators and mean values for both pure and mixed states as described in section 4 and 5. It is also possible to construct the GNS representation in a straight forward manner. We will not go into the details here.

7.2 Generalisation to arbitrary *n*.

The generalisation to arbitrary n is also very straight forward. It is easy to obtain a general expression for the family of idempotents defined in the previous section. They are

$$\varepsilon_{ij} = \frac{1}{n}\sum_{j=0}^{n-1}\omega^{-ji}(e_2)^j \qquad i,j = 0,1,.....n-1. \tag{47}$$

Each one of these idempotents can be used to generate a set of n left ideals so that the techniques discussed in section 4 and 5 can also easily be generalised to arbitrary n.

8. Hilbert Space and the GNS Construction.

Now we want compare our approach with the approach through Hilbert space. This means showing how the minimal left ideals are related to vectors in a Hilbert space. In order to begin we must first introduce the notion of ω-equivalence classes of elements of the algebra, A. To do this consider two elements A and B of A. Whenever $A - B \in L_\omega$, we say A and B are ω-equivalent. Since L_ω is a linear sub-space of A it is clearly an equivalence relation. For every $A \in$ A we denote its equivalent class by $\Omega(A)$.

We now equip the set of equivalent classes with a vector space structure I_ω by defining

$$\lambda\Omega(A) + \mu\Omega(B) = \Omega(\lambda A + \mu B) \qquad \forall\, A, B \in \mathrm{A}, \quad \lambda, \mu \in \Re \text{ or } C$$

We can equip I_ω with a pre-Hilbert space structure by defining

$$(\Omega(A), \Omega(B)) = \Omega(A^* A)$$

With this we can define a norm $|\Omega(A)|^2 = \Omega(A^* A)$. If necessary we can use this norm to complete I_ω thus forming a Hilbert space.

8.1. Examples.

To illustrate the implications of this structure we again examine the example $R_{1,1}$ given above. We take the algebraic spinor defined above in equation (18), namely,

$$\Psi_{L2} = \frac{a}{2}(1 - e_{12}) + \frac{b}{2}(e_1 - e_2).$$

We then see clearly that 1 and $e_{12} \in \Omega(1)$, while e_1 and $e_2 \in \Omega(e_1)$. The Hilbert space is thus two-dimensional so we can write

$$H_\omega = \nu\Omega(1) + \rho\Omega(e_1) \tag{48}$$

Where $\Omega(1)$ and $\Omega(e_1)$ are orthogonal, providing a basis for the two-dimensional Hilbert space H_ω.

Notice that when we pass to the equivalence class we lose some of the mathematical structure in that both $\Psi_{\rho 1}$ and $\Psi_{\rho 2}$ form equivalent Hilbert spaces. In standard quantum mechanics it does not seem necessary to distinguish between these two spaces.

In the case of the nonions we can show that the three ideals I_{L1}, I_{L2} and I_{L3} belong to a single equivalence class. Thus with the obvious generalisation of equation (48) we can construct a Hilbert space by writing

$$H_{\omega 1} = \mu\Omega(1) + \nu\Omega(e_1) + \rho\Omega(e_{11}). \qquad (49)$$

This gives a three-dimensional Hilbert space as expected. Notice that in this case we are dealing with a pure state so that the GNS construction gives us an irreducible representation.

We could generalise to the mixed state by choosing a different state ω'. For example we could start with the density matrix given in equation (36) but choosing for the Hamiltonian $H = -B\pi(e_2)$. The GNS construction will go through exactly as explained in section 5. As the example chosen is not motivated by any relevant physical system we will not discuss it further here but leave it as an exercise for the interested reader.

It should also be stressed that these are not the only left ideals in the algebra. Indeed there are an infinite number of them obtained from inner automorphism. Thus we can always find another left ideal by first finding a new set of idempotents using

$$\varepsilon' = S\varepsilon S^{-1} \qquad S \in \mathrm{A} \qquad (50)$$

It might be of some interest to note that a sub-set of these S generate a set of discrete Fourier transformations.

8.2 Hilbert Space continued.

We now define for all $A \in \mathrm{A}$

$$\pi_\omega(A) : H_\omega \to H_\omega, \qquad \text{so that} \qquad \pi_\omega(A)\Omega(B) = \Omega(AB). \tag{51}$$

We must ensure that this definition implies that $\Omega(AB')$ is independent of B' in $\Omega(B)$ so that it does not matter which Ω-equivalent element B is used to label the equivalence class $\Omega(B)$. To this effect

$$|\Omega(AB) - \Omega(AB')|^2 = \left|\Omega\left(A(B-B')\right)\right|^2 = \Omega\left(\left(A(B-B')\right)^*\left(A(B-B')\right)\right)$$

$$= \Omega\left((B-B')^* A^* A(B-B')\right) \le \|A\|^* . |\Omega(B-B')|^2 = 0$$

This is zero because B–$B' \in I_{\rho\omega}$. Thus it does not matter which element in the equivalence class is used to label the class.

Now we need to verify that $\pi_\omega(A)$ is, indeed, a representation.

$$\pi_\omega(\lambda A + \mu B)\Omega(C) = \Omega\left((\lambda A + \mu B)C\right) = \lambda\Omega(AC) + \mu\Omega(BC)$$

$$= (\lambda\pi_\omega(A) + \mu\pi_\omega(B))\Omega(C)$$

$$\pi_\omega(AB)\Omega(C) = \Omega\left((AB)C\right) = \Omega\left(A(BC)\right) = \pi_\omega(A)\pi_\omega(B)\Omega(C)$$

$$\left(\pi_\omega(A)^*\Omega(C), \Omega(D)\right) = \left(\Omega(C), \pi_\omega(A)\Omega(D)\right) = \omega\left(C^* AD\right)$$

$$= \omega\left((AC)^* D\right) = \left(\Omega\left(A^* C\right), \Omega(D)\right) = \left(\pi_\omega(A^*)\Omega(C), \Omega(D)\right)$$

Thus π_ω is a homorphism of the involutive algebra A to the set of operators acting on the pre-Hilbert space $I_{\rho\omega}$. It is not difficult to show that, if necessary, we can extend $\pi_\omega(A)$ to a bounded operator on H_ω in a unique manner. Furthermore since $\pi_\omega(A^*) = \pi_\omega(A)^*$, $\pi_\omega(A)$ is self-adjoint for every $A \in$ A.

8.3. Return to the Example.

Returning to $R_{1,1}$, in the orthogonal basis $\Omega(1)$, $\Omega(e_1)$ we find

$$\pi_\omega(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \pi_\omega(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\pi_\omega(e_2)=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} \qquad \pi_\omega(e_{12})=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$$

Thus we have produced an irreducible representation and we can write

$$\pi_\omega(\alpha 1+\beta e_1+\gamma e_2+\delta e_{12})=\begin{pmatrix}\alpha+\delta & \beta-\gamma\\ \beta+\gamma & \alpha-\delta\end{pmatrix} \tag{52}$$

while

$$\Omega(1)=\begin{pmatrix}1\\ 0\end{pmatrix} \quad \text{and} \quad \Omega(e_1)=\begin{pmatrix}0\\ 1\end{pmatrix} \tag{53}$$

Thus in a pure state defined by $\omega_1(A)$ given above, we find that the representation for Ω is the familiar real Pauli spinor.

In the case of the nonions equation (53) is generalised to

$$\Omega(1)=\begin{pmatrix}1\\ 0\\ 0\end{pmatrix} \qquad \Omega(e_1)=\begin{pmatrix}0\\ 1\\ 0\end{pmatrix} \text{ and } \qquad \Omega(e_{11})=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}$$

Giving us the basis for the three-dimension representation discussed in section 8.1. It is not difficult to show that the corresponding matrix representation can be written as

$$\pi_{\omega_1}(e_1)=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix};\ \pi_{\omega_1}(e_2)=\begin{pmatrix}1 & 0 & 0\\ 0 & \omega & 0\\ 0 & 0 & \omega^2\end{pmatrix} \text{and } \pi_{\omega_1}(e_{12})=\begin{pmatrix}0 & \omega & 0\\ 0 & 0 & \omega^2\\ 1 & 0 & 0\end{pmatrix}$$

The rest of the matrixes can be found using the multiplication table above.

Notice also that this approach enables us to attribute to each $A \in \mathbb{A}$, a vector

$$\Omega(A)=\pi_\omega(A)\Omega(1) .$$

When generalising to include an infinite algebra we require the representation space H_ω to be defined in terms of a strong closure of these vectors. This means that $\Omega(1)$ is cyclic for π_ω. A vector is cyclic relative to the algebra if the set $\{\pi_\omega(A)\Omega \mid A \in \mathrm{A}\}$ is dense in H_ω. This means that for any vector $\psi \in \mathcal{H}_\omega$ $\|\psi - \pi_\omega(A)\|$ can be made as small as desired.

In conclusion then our construction associates to every state ω on A a cyclic representation π_ω of A. We can then write the expectation value as

$$\omega(A) = \langle\Omega|\pi_\omega(A)|\Omega\rangle$$

where we have written $\Omega(1) = |\Omega\rangle$.

8.4 The Limit n $\to\infty$

The simplest way to proceed to generalise the n-ions is to go to the limit $n \to \infty$. This limit has already been discussed by Hiley and Monk (1993) and Hiley (2001), their argument being based on earlier work by Weyl (1931). We will briefly outline the argument again here for the sake of completeness.

We have written $e_1 = \exp[i\xi P]$ and $e_2 = \exp[i\eta X]$ where $\xi = 2\pi/\delta pn$ and $\eta = 2\pi/\delta xn$ so that we can regard e_1 as a translation in space and e_2 as a translation in momentum space. Thus we can regard $(e_1)^{-s}$ as translating from $x_k \to x_{k-s}$. In the limit this corresponds to the transformation $\psi(x) \to \psi(x-s)$. Similarly $(e_2)^t$ takes $x_k \to \omega^{kt}x_k$, which in the limit corresponds to $\psi(x) \to \exp[itx].\psi(x)$. This shows how the Schrödinger representation emerges from the discrete Weyl algebra. In this case the discrete Fourier transformation mentioned above in section 7.1 becomes the usual Fourier transformation that takes the x-representation to the p-representation.

If we now examine the limiting process on the primitive idempotent ε_{00} we find

$$\varepsilon_{00}(n) = \frac{1}{n}\sum_\beta \exp[i\beta X] \;\to\; \frac{1}{2\pi}\int d\beta \exp[i\beta X] \;\to\; \delta(x) \qquad (53)$$

Thus we see that in the limit the Weyl idempotent $\varepsilon_0(n)$ plays the role of the Dirac delta function in the continuum algebra. If we start directly from the Heisenberg algebra defined through the standard commutation relations $[X, P] = i$, we miss the idempotent,

the Dirac delta function which we must add later when we have constructed the Hilbert space.

If we remain in the algebraic structure then we can formally write $\varepsilon_{00}(n) = \Delta$ then we find the relations

$$X\Delta = 0, \qquad \Delta P = 0, \qquad \text{and} \qquad \Delta^2 = \Delta \tag{54}$$

If we identify $E = \Delta\dagger$ introduced by Frescura and Hiley (1984) then the following relation result:

$$EX = 0, \qquad PE = 0, \qquad \text{and} \qquad E^2 = E \tag{56}$$

The set $\{I, X, P, E\}$ then generates what we have called the extended Heisenberg algebra. Further details of this structure can be found in Hiley (2001).

9 Conclusions.

We have shown how an approach to quantum mechanics can be built from the algebraic structure of the Clifford algebra and the discrete Weyl algebra (or the generalised Clifford algebra). These algebras can be treated by the same techniques that do not require Hilbert space yet enable us to calculating mean values required in quantum mechanics. In the appendix we compare the two approaches in table form. We have also shown how these techniques are related to the Hilbert space, which effectively requires additional structure that involves the space-time continuum. Yet these algebras already contain geometric feature not only of space-time itself, but of a generalised phase space as already pointed out in Bohm and Hiley (1981 and 1983). The full implications of this geometric structure will be discussed in another publication.

10. Appendix.

In this appendix we list the correspondence between the algebraic approach to quantum mechanics which we call "The Algebra of Process" and the usual approach which we call "The Algebra of Observables". We also include the extra step which involves a projection of the ket vector space into $L^2(x)$.

Algebra of Process.

*Algebra **A**

Left ideal, $L \in \mathfrak{l}_L$

Generated by idempotents ε_j,
$\varepsilon_j^2 = \varepsilon_j \rightarrow A\varepsilon_j$.

Basis of ideal ε_{ij} with $\varepsilon_j \rightarrow \varepsilon_{ij}$
$[1 \geq i \geq n$, AND $1 \geq j \geq n]$

Right ideal, $R \in \mathfrak{l}_R$

Generated by idempotents ε_j
$\varepsilon_j^2 = \varepsilon_j \rightarrow \varepsilon_j A$.

$[*: L \rightarrow R,\ \varepsilon_i^* = \varepsilon_i\ ]$

Inner product. $\omega(RL) = \langle R|\, L\rangle$

$\omega : \varepsilon_{ij} \rightarrow \delta_{ij},\ \forall\ \varepsilon_{ij}$,

[Trace : coeff of λ^{r-1} of characteristic poly.]

$\omega(L^*L) = \omega(\varepsilon_i A \varepsilon_i) = \omega(\lambda_A \varepsilon_i) = \lambda_A \in R$

Outer product, $L.R \rightarrow B\varepsilon_{ii}C$
$L.L^* \rightarrow B\varepsilon_{ii}B^*$

Completeness relation.

$$1 = \sum_i \varepsilon_{ii}$$

Algebra of Observables.

† Algebra **A**

External vector space, $|\ \rangle \in V$
$\Rightarrow \langle x|\, \psi\rangle = \psi(x)$

Basis $|\, i\,\rangle$
$[1 \geq i \geq n]$

Dual vector space, $\langle|\ \in V^*$
$\Rightarrow \langle \psi|\, x\rangle = \psi^*(x)$

$[\dagger : |\ \rangle \rightarrow \langle|\ ]$

Inner product $\langle|\ \rangle$
$\Rightarrow \int \psi^*(x)\psi(x)\, d^3x$

Outer product, $|\ \rangle\langle|$
↓
Projection operator $P_i = |\, i\,\rangle\langle i|$

Completeness relation.

$$1 = \sum_i |i\rangle\langle i|$$
$$\Rightarrow \sum_i \langle \psi_i|x'\rangle\langle x|\psi_i\rangle = \delta(x' - x)$$

L ensity operator, $\rho = \sum_i B\varepsilon_{ii}B^*$

$$\rho = \sum_i \mu_{ii}\varepsilon_{ii} + \sum_{i,j(i<j)} \mu_{ij}\left(\varepsilon_{ij} + \varepsilon_{ji}\right)$$

[ρ =ρ†, Tr(ρ) = 1, ρ+ve and linear]

For diagonal ρ

$$\rho = \sum_i \lambda_i \varepsilon_{ii}$$

Expectation values.

$$\langle\ \rangle = \omega(\rho A) = \left(\sum_i \varepsilon_{ii} B^* AB\varepsilon_{ii}\right)$$

GNS construction

$\rho = D^*D = D^*ID$, i.e. $\exists\ D = \sum_i \sqrt{\lambda_i}\varepsilon_{ii}$

$$\langle A\rangle = \omega\left(D^* AD\right)$$

Density operator, ρ,

$$\rho = \sum_i \mu_{ii}|\psi_i\rangle\langle\psi_i| + \sum_{i,j(i<j)} \mu_{ij}\left(|\psi_i\rangle\langle\psi_j| + |\psi_j\rangle\langle\psi_i|\right)$$

For diagonal ρ

$$\rho = \sum_i \lambda_i |i\rangle\langle i|$$

$$\rho(x', x) = \sum_i \lambda_i \psi^*(x')\psi(x)$$

Expectation values.

$$\langle A\rangle = Tr(\rho A) = \sum_r \sum_i \langle r|\lambda_i A|i\rangle\langle i|r\rangle = \sum_r \langle r|A|r\rangle$$

$$\Rightarrow \langle A\rangle = \sum_i \lambda_i \int \psi_i^*(x)\psi_i(x)d^3x$$

GNS construction

$\exists\,|\,\Omega\,\rangle$ a cyclic vector

$$\langle A\rangle = \langle\Omega|\pi(A)|\Omega\rangle$$

11. References.

I. M. Benn and R. W. Tucker, (1983), *Fermions without Spinors*, Comm. Math. Phys. 8 ', 341-362.

G. Birkhoff (1966), Lattice Theory, AM. Math. Soc. Coll. Pub. **25**.

D. Bohm and B. J. Hiley, (1981), *On a Quantum Algebraic Approach to a Generalised Phase Space*, Found. Phys., **11**, 179-203.

D. Bohm and B. J. Hiley, (1983), *Relativistic Phase Space Arising out of the Dirac Algebra*, in Old and New Questions in Physics and Theoretical Biology, Ed., A. van der Merwe, pp. 67-76, Plenum, New York.

O. Bratteli and D. W. Robinson, (1979), *Operator Algebras and Quantum Statistical Mechanics I and II*, Springer, Berlin.

M. R. Brown, and B. J. Hiley, (2000), *Schrödinger Revisited: an Algebraic Approach.* quant-ph/0005026.

E. Celeghini, M. Rasetti, and G. Vitiello, (1992), *Quantum Dissipation*, Ann. Phys. 215, 156-170.

A. Crumeyrolle, (1990), *Orthogonal and Symplectic Clifford Algebras: Spinor Structures*, Kluwer, Dordrecht.

P. A. M. Dirac, (1965), *Quantum Electrodynamics without Dead Wood*, Phys. Rev., **139B,** 684-90.

G. G. Emch, (1972), *Algebraic Methods in Statistical Mechanics and Quantum Field Theory*, Wiley-Interscience, New York,

F. A. M. Frescura and B. J. Hiley, (1978), *Minimum and Characteristic Polynomials and Non-relativistic Spinors in the Pauli Clifford Algebra*, (unpublished).

F. A. M. Frescura and B. J. Hiley, (1980) *The Implicate Order, Algebras, and the Spinor*, Found. Phys., **10**, 7-31.

F. A. M. Frescura and B. J. Hiley, (1980) *The Algebraization of Quantum Mechanics and the Implicate Order*, Found. Phys., **10**, 705-22.

F. A. M. Frescura and B. J. Hiley, (1984), *Algebras, Quantum Theory and Pre-space, Rev. Brasil Fis.*, Volume Especial, Os 70 anos de Mario Schonberg, 49-86.

F.A.M. Frescura and B. J. Hiley, (1987), *Some Spinor implications unfolded, in Quantum Implications:* Essays in Honour of David Bohm, ed B. J. Hiley and F. D. Peat, (op. cit. 50) pp. 278-288,

R. Haag, (1992), *Local Quantum Physics*, Springer, Berlin.

B. J. Hiley, (1991), *Vacuum or Holomovement*, in The Philosophy of Vacuum, ed., S. Saunders and H. R. Brown, pp. 217-249, Clarendon Press, Oxford.

B. J. Hiley, and N. A. M. Monk, (1993), *Quantum Phase Space and the Discrete Weyl Algebra*, Mod. Phys. Lett., **A8**, 3225-33.

B. J. Hiley, (1995), *The Algebra of Process,* in Consciousness at the Crossroads of Philosophy and Cognative Science: Proc. of the Tempus Project "Phenomenology and Cognitive Science" Maribor, Slovenia, 23-7 August, 1994, ed., B. Borstner and J. Shawe-Taylor, pp. 52-67, Imprint Academic, Thorverton.

B. J. Hiley and M. Fernandes, (1997), *Process and Time*, in Time, Temporality and Now, ed., H. Atmanspacher and E. Ruhnau, 365-383, Springer-Verlag.

B. J. Hiley, (2001), *Non-Commutative Geometry, the Bohm Interpretation and the Mind-Matter Relationship*, Computing Anticipatory Systems: CASYS 2000-Fourth International Conference. edited by D. M. Dubois, pp. 77-88, AIP.

B. J. Hiley, (2001) *A Note on the Role of Idempotents in the Extended Heisenberg Algebra*, ANPA.

L. Kauffman, (1991), *Knots and Physics*, pp. 381-402, World Scientific, Singapore.

N. A. M. Monk and B. J. Hiley, (1998), *A unified algebraic approach to quantum theory*, Found. Phys. Letts, **11**, 371-377.

A. O. Morris, (1967) *On a Generalised Clifford Algebra*, Quart. J. Math. **18**, 7-12.

I. Progigne, (1980), *From Being to Becoming: Time and complexity in the physical sciences,* W. H, Freeman, San Francisco,

I. Prigogine, *Time's Arrow and the Laws of Nature*, First Draft April 26, 1994.

M. Rédei (1996), *Why von Neumann did not like the Hilbert space formalism of quantum mechanics (and what he like instead)*, Stud. Hist. Phil. Mod. Phys. **27**, 493-510.

M. Schönberg, (1957) Nuovo Cimento, Supp. **VI**, 356-380.

J. J. Sylvester (1882) John Hopkins Uni. Circulars, **I**, p. 241, **III**, p. 7.
These are reproduced in J. J. Sylvester, *Collected Mathematical Papers of James Joseph Sylvester*, **III**, p. 647 and **IV**, p. 122, AMS, Chelsea Publishers. (1973),

H. Weyl, (1931), *The Theory of Groups and Quantum Mechanics*, Dover, London.

A New View of Sommerfeld's Fine Structure Constant And Its Consequences

PETER MÖBIUS

Postbox 200235 / D - 01192
Dresden / Germany

Abstract

Sommerfeld's fine structure constant is decomposed into 2 universal electrical resistancies, the impedance of the vacuum and the quantum Hall resistance. With these quantities a new quantization procedure is postulated, leading to new special constants of unconventional dimension (SCUD's) having the same value for the classical and assigned quantum system. Extending these ideas to the case of gravitational interaction mass relations are deduced, leading to "mass analogues" With this procedure an instructive parametrization of the masses of the heavy leptons and the heavy gauge vector bosons is derived, using an expansion with respect to power series of the inverse square root of Sommerfelds fine structure constant, having a reasonable agreement with the experimental values.

Keywords Sommerfeld's fine structure constant, energy spectra, mass relations

1 Introduction to important constants of nature

For the theoretical description of different phenomena in nature the fundamental constants of nature play an essential role. They should be a combination of a numerical value with a physical dimension, expressed in certain units. For the description of classical physics the essential constants are known up to the beginning of the 20^{th} century. But with the beginning of atomic, nuclear and high energy physics a considerable number of new constants appeared, where Plancks action constant played an essential role.
Let us start with the constants of classical physics [1]

speed of light	c	$= 299\,792\,458$	ms^{-1}
Newtons gravitational constant	G_N	$= 6.6725985 \cdot 10^{-11}$	$m^3kg^{-1}\,s^{-2}$
permittivity of free space	ε_0	$= 8.854\,187\,817 \cdot 10^{-12}$	Fm^{-1}
permeability of free space	μ_0	$= 12.566\,370\,614\,10^{-7}$	NA^{-2},

the last 2 constants are connected with the speed of light by $c^{-2} = \varepsilon_0\mu_0$,
impedance of free space $Z = \sqrt{\mu_0}/\varepsilon_0 = 1/\varepsilon_0 c = 376,72\,\Omega$.
Going over to quantum theory and elementary particle physics
an important role play

Plancks action konstant	h	$= 6.626075 \cdot 10^{-34}$	Ws^2
electric charge of the electron	$\|e\|$	$= 1.602\,177 \cdot 10^{-19}$	C
mass of the electron	m_e	$= 0.510\,999$	MeV/c^2
mass of the proton	m_p	$= 938.272\,313$	$MeV/c^2 = 9.109 \cdot 10^{-31}kg$
Bohr magneton	μ_B	$= 5.788\,382 \cdot 10^{-11}$	$MeVT^{-1}$.

So far no attempt is made to derive the numerical values of these constants.
An important role for the description of the interaction of electromagnetic radiation with electric charge plays Sommerfelds fine structure constant given by

$$\alpha = e^2/4\pi\varepsilon_0\hbar c, \quad \alpha = 1/137.035\,989, \tag{1}$$

which is dimensionless, meaning that its value is independent in which units e, h, c are measured. How can this fact be explained? The first attempt is to assume that it is the quotient of 2 physical quantities of the same dimension. So we try the ansatz

$$\alpha = Z/R_0 \qquad \text{with} \quad Z = 1/\varepsilon_0 c \quad \text{and} \quad R_0 = 2h/e^2, \tag{2}$$

where Z is the above given impedance of the free space $Z = 376,72\,\Omega$ and R_0 a so far unknown quantity of the size

$$R_0 = 51,625\ \text{k}\Omega. \tag{3}$$

How could such a big resistance be explained up to 1985? So far only in the theory of electric circuits and filters resistancies of such order occured, but why should there appear Plancks action quantum h? No useful explanations were found up to 1985.

2 Nanoelectronics

With increasing integration of the density of devices on chips a transition took place from microelectronics to nanoelectronics. While microelectronics is dominated by classical physics, in nanoelectronics (1nm $= 10^{-9}$m) quantum effects can occur. To compare it with atomic physics one can roughly say, that 1nm corresponds to $20 \cdot a_1$, where a_1 is the first Bohr radius in the hydrogen atom, given by $a_1 = \varepsilon_0 h^2/\pi\, m_e\, e^2 \approx 0,53 \cdot 10^{-10}$m.

Because electrons play an essential role, it is instructive to look for the action quantum per unit charge h/e = 4,132Vs,
permitting to say, that it is related to a special flux quantum measured in Vs. Now the Josephson effect provided that the elementary flux quantum is given by $\Phi_0 = h/2e = 2,066 \cdot 10^{-15}$Vs. (Giaver, Asaki, Josephson Nobel prize 1973)
(Historical remark: when lecturing about the Josephson effect I suggested already, that the quantities $(h, e, 2)$ are so important, that at least another Nobel prize containing these 3 quanitities $(h, e, 2)$ will follow).
More than ten years later the "Quantum-Hall-effect" was discovered, providing for the quantum-Hall-resistance the relation

$$R_H = h/ne^2 \approx 1/n\,(25,813)\,\mathrm{k\Omega}\,,\ (n = 1, 2, 3\ldots) \tag{4}$$

and Klaus v. Klitzing received the Nobel prize 1985 for it. The essential idea is now that the conductivity occurs in steps of $e^2/h = 38,740\,\mu$S. Soon followed the triangle relation for (U, I, f) for the "1-electron-transistor" containing voltage, current and frequency (Fig. 1).

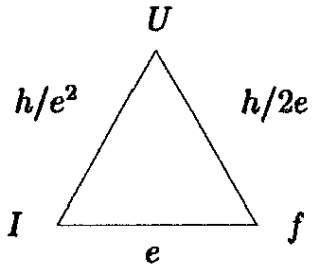


Figure 1: The triangle relation for (U, I, f)

The corresponding relations between these quantities are now

$$U = (h/e^2)I, \qquad U = (h/2e)f \qquad \text{and } I = ef. \tag{5}$$

Now we can understand the quantity $R_0 = 2R_H(n = 1) = 51,625\,\mathrm{k\Omega}$ and find the connection to Sommerfelds fine structure constant

$$\alpha = e^2/4\pi\varepsilon_0\hbar c = \frac{1}{2}\cdot\frac{1}{\varepsilon_0 c}\cdot\frac{1}{h/e^2} = \frac{1}{2}\left(\frac{Z}{R_H(1)}\right) = Z/2R. \tag{6}$$

Let us finish this chapter with an instructive relation characterizing the type of nanotechnology by the power, which is necessary for the use of one chip (usually with 30 cm diameter) having N_b bits. It is given by

$$P = \varepsilon_b N_b f \tag{7}$$

where ε_b is the energy needed to create 1 bit. This quantity ε_b characterizes the applied technology. In the case of Silicium-technology it is approximately given by $\varepsilon_b(Si) = 10^{-13}$Ws/bit, but there exist other ones, e.g. RSFQ-technology needing only 10^{-18}Ws/bit. The quantity N_b is the number of bits of the chip. For a diameter of 30 cm possible values are $N_b(I) = 1$Mbit to $N_b(II) = 1$Gbit .
Finally f is the applied frequency, which is nowadays between 1 MHz and 1 GHz. Let us look at 2 different examples I and II:

$$N_b(I) = 1\,\text{Mbit}, f(I) = 1\,\text{MHz} \quad \text{and} \quad N_b(II) = 1\,\text{Gbit}, \; f(II) = 1\,\text{GHz}, \quad (8)$$

$$P(I) = \varepsilon_b(Si)\, N_b(I)\; f(I) \; = 10^{-13}\text{Ws/bit}\; 10^6\text{bit}10^6\text{s}^{-1} = 0,1\,\text{W}, \quad (9)$$

$$P(II) = \varepsilon_b(Si)\, N_b(II)\, f(II) = 10^{-13}\text{Ws/bit}\; 10^9\text{bit}10^9\text{s}^{-1} = 100\,\text{kW}. \quad (10)$$

This power is mainly needed for cooling. It is worthwhile to compare it with the power of a gas stove of a 1-family house in Dresden, having in average 20 kW.

3 New Quantization Rules

The decomposition of Sommerfelds fine structure constant into a "classical" part Z and a "quantum theoretical part" R leads to a new quantization procedure revealing two new phenomena

- SCUD's : special constants of unconventional dimension
- Mass analogues: mass relations for heavy leptons and heavy gauge vector bosons with respect to reference particles.

Let us start with classical quantities C_n and the corresponding quantum theoretical ones Q_{np}, indicating by p that there might be different ones for a given n. The new quantization procedure has the form

$$C_n f_{1p}(Z) = Q_{np}(R)\; f_{2p}(R), \quad n = 1 \cdots N, \quad (11)$$

where Z and R are given above and n characterizes the number the different classical quantities (lengthes, energies ...), and p, characterizing the power of Z and R, may take integer and noninteger values

$$f_{ip}(X) = X_i^p, \; i = 1,2 \quad \text{and} \quad X_1 = k_1 Z, \; X_2 = k_2 R, \quad (12)$$

with k_i having so far only the values $2^1, 2^0, 2^{-1}$.

This relation is the backbone of the following considerations, where now numerous applications in different fields of physics, reaching from atomic to gravitational and high energy physics will be investigated. To classical systems C_n and quantum systems Q_{np} new constants S_n^p can be assigned having the same numerical value in both systems, if they are combined with an unconventional dimension

$$S_n^p = \mathrm{C_n\, f_{1p}(Z) = Q_{np}\, f_{2p}\,(R)}\,. \quad (13)$$

These constants S_n^p are called SCUD'S, which are so far not treated in the literature.

4 Application to Atomic Physics

Let us start with a basic quantity of elementary particle physics, which is nowadays surprisingly the classical electron radius [1] defined by the relation

$$mc^2 = \frac{e^2}{4\pi\varepsilon_0 r_e}, \; \text{giving}\; r_e = \frac{e^2}{4\pi\varepsilon_0 mc^2} = 2,81794 fm. \quad (14)$$

Using now r_e as C_1 and putting $p=-1$ in (11) used in the form

$$r_e f_{1,-1}(Z) = Q_{1,-1} f_{2,-1}(2R) \quad \text{giving} \quad \frac{r_e}{Z} = \frac{Q_{1,-1}}{2R} \tag{15}$$

and with the help of $Z = \sqrt{\frac{\mu_0}{\varepsilon_0}} = \frac{1}{\varepsilon_0 c}$,

$$\text{we obtain } \frac{e^2 \varepsilon_0 c}{4\pi\varepsilon_0 mc^2} = Q_{1,-1}\frac{e^2}{2h} \quad \text{meaning} \quad Q_{1,-1} = \frac{\hbar}{mc} = \bar{\lambda}_e = 386,159 fm. \tag{16}$$

So to the classical electron radius we assign with the help of (15) the Compton wavelength of the electron. Now due to the relation (15) we get a new physical quantity

$$S_1^{(1)} = \frac{r_e}{Z} = \frac{\bar{\lambda}}{2R},\ S_1^{(1)} = \frac{e^2}{4\pi mc},\ S_1^{(1)} = 7,84\,\mathrm{am}\Omega^{-1} (1\mathrm{am} = 10^{-18}\mathrm{m}). \tag{17}$$

This is a new constant, having an unconventional physical dimension $[\mathrm{m}\Omega^{-1}]$, therefore it is called 'SCUD', meaning special constant of dimension. It is now one aim of this paper to derive numerous SCUD'S in different field of physics and to discuss its consequences. They are 'invariant' by the transition from the classical to the quantum system, so they form some common 'roof' of classical and quantum theory.
In the next case we use again $C_1 = r_e$, but put $p=-2$ providing

$$r_e f_{1,-2}(Z) = Q_{1,-2} f_{2,-2}(2R) \quad \text{or} \quad \frac{e^2(\varepsilon_0 c)^2}{4\pi\varepsilon_0 mc^2} = Q_{1,-2}\frac{e^4}{4h^2}, \tag{18}$$

$$\text{meaning} \quad Q_{1,-2} = \frac{\varepsilon_0 h^2}{\pi m e^2} = a_B \quad \text{with} \quad a_B = 52,9\,\mathrm{pm}, \tag{19}$$

where a_B is the first Bohr radius of the hydrogen atom in the case of a nucleus with an infinite mass.
The corresponding SCUD $S_1^{(2)}$ has the value

$$S_1^{(2)} = \frac{e^2\varepsilon_0 c}{4\pi\varepsilon_0 mc},\quad S_1^{(2)} = 1,986 \cdot 10^{-20}\,\mathrm{m}\Omega^{-2}. \tag{20}$$

Now let us pass to the case using again C_1 but with $p=-3$ giving

$$\frac{C_1}{Z^3} = \frac{Q_{1,-3}}{(2R)^3} \quad \text{leading to} \quad Q_{1,-3} = \frac{2\varepsilon_0^2 ch^2}{\pi m e^4} = a_N \text{ with } a_N = 7,25\,\mathrm{nm}. \tag{21}$$

This is a size typical in nanophysics, so we conjecture now that a_N represents some critical thickness of layers in systems of nanoelectronics [2].

The corresponding SCUD has the value

$$S_1^{(3)} = \frac{e^2(\varepsilon_0 c)^2}{4\pi mc}, \quad S_1^{(3)} = 5,272 \cdot 10^{-23}\mathrm{m}\Omega^{-3}, \tag{22}$$

which is a very small amount combined with an unconventional dimension. Now it is instructive to give a survey for the three SCUD'S connected with r_e, having the shape

$$S_1^{(1)} = \frac{e^2}{4\pi mc}, \quad S_1^{(2)} = \frac{e^2\varepsilon_0 c}{4\pi mc}, \quad S_1^{(3)} = \frac{e^2(\varepsilon_0 c)^2}{4\pi mc}, \tag{23}$$

which in a short hand notation can be written in this case

$$S_1^{\nu} = S_1^{(1)} \cdot Z^{1-\nu}, \quad \nu = 1,2,3. \tag{24}$$

This demonstrates that this new quantities have certain discrete values in the set of the SCUD'S, all can be expressed by only classical quantities, though they are related with classical and quantum objects. Is it some new kind of 'discretization' of classical objects, meaning that these classical objects can have only certain specific values? Do we get a new view for classifying objects in atomic and nanophysics, or even high energy physics?

5 Treatment of the Energy Levels of the Hydrogen Atom

As a classical quantity we use the rest energy of the electron, divided by a factor of 2 for convenience. So we begin with $C_2 = \frac{1}{2}mc^2$ and $p = 2$ and the relation (11) in the form

$$C_2 f_{1,2}(Z) = Q_{2,2} f_{2,2}(2R), \quad \text{giving} \quad Q_{2,2} = \frac{me^4}{8\varepsilon_0^2 h^2} = E_H, \quad E_H = 13,605\,\mathrm{eV}, \tag{25}$$

the binding energy of the hydrogen atom in the case of a nucleus with infinite mass. The corresponding SCUD $S_2^{(1)} = C_2 f_{1,2}(Z)$ is given by

$$S_2^{(1)} = \frac{mc^2}{2(\varepsilon_0 c)^2} \quad \text{with the value} \quad S_2^{(1)} = 36,255\,\mathrm{GeV}\Omega^2. \tag{26}$$

In the next step we look at the fine structure splitting in the hydrogen atom, i.e. the energy difference between the two excited levels with principal quantum number $n = 2$ but with different spin states $j = \frac{3}{2}$ and $j = \frac{1}{2}$, called $\Delta E_F(2)$, which is of the order $\Delta E_F(2) = 45,3\,\mu\mathrm{eV}$. Here it is appropriate to use the relation (11) with the same C_2 but $p = 4$ in the form

$$C_2 f_{1,4}\left(\frac{Z}{2}\right) = Q_{2,4} f_{2,4}(2R), \quad \text{giving} \quad \frac{1}{2}mc^2\left(\frac{Z}{2}\right)^4 = \Delta E_F(2)(2R)^4. \tag{27}$$

Leading to

$$\Delta E_F(2) = \frac{1}{32} mc^2 \left(\frac{Z}{2R}\right)^4 \quad \text{or} \quad \Delta E_F(2) = E_H \frac{\alpha^2}{16}, \tag{28}$$

$$\text{where the relation} \qquad \frac{1}{2} mc^2 = \frac{E_H}{\alpha^2} \quad \text{is used}. \tag{29}$$

This result can easily be derived with the help of the Dirac equation [3].
Now the corresponding quantity $S_2^{(2)}$ is given by

$$S_2^{(2)} = \frac{mc^2}{32(\varepsilon_0 c)^4} \quad \text{with the value} \quad S_2^{(2)} = 321,8\,\text{TeV}\Omega^4, \tag{30}$$

which we can assign to the fine structure splitting in the hydrogen atom for $n = 2$. In the hydrogen atom the number of states belonging to the shell with principle quantum number n is equal to $2n^2$. In the case of $n = 2$ these are 8 states, a quartet belonging to $j = \frac{3}{2}$ and two doublets having $j = \frac{1}{2}$. The energy difference between the two doublets is just the famous Lambshift [3] with the splitting of $\Delta E_L(2) \approx 4,37\mu\text{eV}$. Now we shall try to get some approach to this effect by starting again with C_2 but taking $p = 5$ in (11) in the form

$$C_2 f_{1,5}(Z) = Q_{2,5} f_{2,5}(2R), \quad \text{providing} \quad Q_{2,5} = \frac{1}{2} mc^2 \left(\frac{Z}{2R}\right)^5 = E_H \alpha^3, \tag{31}$$

where the relation (29) has been used again. To a good approximation the Lambshift [4] is given by the approximation

$$\Delta E_L(2) \approx \sqrt{\frac{2}{3}} E_H \alpha^3, \quad \text{giving the value} \quad \Delta E_L(2) \approx 4.320\,\mu\text{eV}, \tag{32}$$

which deviates from the standard value of $4,3737\mu\text{eV}$ roughly by 1 % .
Relation (32) indicates that the quantity $Q_{2,5}$ does not agree with a directly observable quantity of the hydrogen spectrum but deviates a factor of $\sqrt{\frac{3}{2}}$ from it

$$Q_{2,5} = \sqrt{\frac{3}{2}} \Delta E_L(2), \tag{33}$$

which is however of the order $0(1) \sim 1,22$, which is an acceptable fact

$$C_2 f_{1,5}(Z) = \sqrt{\frac{3}{2}} \Delta E_L(2) f_{2,5}(2R). \tag{34}$$

Tolerating this we look now at the SQUD $S_2^{(3)}$, considering both sides of (34) separately and using again (29).

Left hand side:

$$C_2 f_{1,5}(Z) = \frac{1}{2} mc^2 Z^5 = \frac{mc^2}{2(\varepsilon_0 c)^5} .$$

Right hand side:

$$Q_{2,5} f_{2,5}(2R) = \sqrt{\frac{3}{2}} \Delta E_L(2)(2R)^5 = \sqrt{\frac{3}{2}}\sqrt{\frac{2}{3}} E_H \alpha^3 (2R)^5 = \frac{1}{2} mc^2 \alpha^5 (2R)^5 =$$
$$= \frac{1}{2} mc^2 Z^5 = \frac{mc^2}{2(\varepsilon_0 c)^5} ,$$

providing $S_2^{(3)} = \frac{mc^2}{2(\varepsilon_0 c)^5}$ with the value

$$S_2^{(3)} = 1,91\,\mathrm{EeV}\Omega^5 \;(1\mathrm{EeV} = 10^{18}\mathrm{eV}) , \tag{35}$$

where the 'energy part' of this constant is extremely huge even in high energy physics.

6 Derivation of Mass Relations for Elementary Particles

When the high energy accelerators started working in the early sixties a number of new objects were discovered, called "elementary particles", though they are not at all elementary. They are divided into leptons, hadrons (baryons and mesons) and heavy gauge vector bosons. In addition there is a "monster particle" called Higgs particle.

Up to now no generally excepted way exists to determine the corresponding mass values. A first attempt was Gell-Mann's "Eightfold way", where the particles were put into SU(3) octets and decuplets. This is however doubtful, because the masses of the particles within one multiplet can differ by a factor of 3. To take care of this, a symmetry-breaking is introduced leading to the Gell-Mann-Okubo mass formulaes [1]

$$\frac{1}{2}(m_N + m_\Xi) = \frac{1}{4}(3m_\Delta + m_\Sigma) , \quad 4m_K^2 = 3m_\eta^2 + m_\pi^2 , \tag{36}$$

where the first one is concerned with the $(\frac{1}{2})^+$ octet of baryons and the other one with the O^- octet of mesons (Gell-Mann, Nobel prize 1969).

Now we shall derive mass relations by an expansion with respect to certain coupling constants. How do we get such an expansion? The fundamental mass of quantum theory is the Planck mass M_p, given by

$$M_p = \sqrt{\frac{\hbar c}{G_N}} , \; M_p = 2,176\,10^{-8}\,\mathrm{kg} \quad \text{or} \quad \mathrm{M_p c^2} = 1,22\,10^{19}\,\mathrm{GeV} . \tag{37}$$

How is it derived? We plot two typical lengthes, the quantum theoretical Compton wave length $\lambda \equiv h/mc$ and the classical Schwarzschild radius $R_s = 2\,G_N \mathrm{m/c^2}$

as function of the mass and ask for the intersection point $\lambda = R_s$ providing $M_p = \sqrt{\frac{hc}{2G_N}}$. Discarding a factor of 2 we get the above result for M_p, having a meaning for chemists, because its value is also $M_p = 21,76\,\mu g$.
Computing the Compton wave length of M_p provides the Planck length

$$L_p = \sqrt{\frac{hG_N}{c^3}}\,,\; L_p = 1,6 \cdot 10^{-35}\,\mathrm{m}\,,$$

which is far beyond all experimental facilities. Now L_p is the basis of the "THEORY of EXTENDED OBJECTS" called "strings".
Now we go the opposite way and try to assign to L_p and M_p the corresponding classical quantities with the help of relation (11) with $p = -1/2$ providing

$$l_{cl} = \frac{e}{c^2}\sqrt{\frac{G_N}{4\pi\varepsilon_0}} \approx 10^{-36}\,\mathrm{m}\,,\; m_{cl} = |e|/\sqrt{4\pi\varepsilon_0 G_N} = 1,86\,\mu g\,. \tag{38}$$

The classical mass can also be derived by the requirement that for it the gravitational force is equal to the electric force between 2 elementary chargese of opposite sign for arbitrary distance

$$G_N \frac{m_{cl}^2}{r^2} = \frac{1}{4\pi\varepsilon_0} \cdot \frac{|e|^2}{r^2}\,.$$

Now we obtain the fundamental relation

$$M_p/m_a = 1/\sqrt{\alpha} \approx 11.7 \gg 1\,, \tag{39}$$

which is in contrast to the relation that the binding energy of the electron in the hydrogen atom E_H is very much smaller than its classical energy $E_{cl} = mc^2$

$$E_H/E_{cl} = \frac{1}{2}\alpha^2 \approx 2 \cdot 10^{-5} \ll 1\,.$$

So we can assign to a classical mass m_{cl} a higher mass m_A (mass analogue) via $m_{cl} \to m_A = m_{cl}/\sqrt{\alpha}$ or an expansion with few terms in the form

$$m \to m_A = m/\sqrt{\alpha}\,(a_0 + a_1/\sqrt{\alpha} + \ldots) \qquad \text{"mass analogue"}\,. \tag{40}$$

Now we apply this idea to the masses of the heavy leptons with masses $m_\mu = 105,6\,\mathrm{MeV/c^2}$ and $m_\tau = 1777,0\,\mathrm{MeV/c^2}$, where the müon and the tauon belong to the 2nd and 3rd generation $g(\mu) = 2$, $g(\tau) = 3$ of leptons. Analogous to the hydrogen atom $C = mc^2/2$, we postulate that $C(\mu)$ and $C(\tau)$ are given by the mass analogues, characterized by g, giving

$$m_L(g) = \frac{2m_e}{\sqrt{\alpha}^{g-1}}(1/\sqrt{\alpha} + \beta(g))\,, \tag{41}$$

with the generation function $\beta(g)$ having the form $\beta(g) = (-1)^{g-1}N_L - 1$, where $N_L = 2$ is the number of generations of heavy leptons, giving $\beta(2) = -3$ and $\beta(3) = 1$. So the expansion with only 2 terms provides already the mass values

$$m_\mu c^2 = \frac{2m_e c^2}{\sqrt{\alpha}}(1/\sqrt{\alpha} - 3) = 104,2\,\mathrm{MeV} \quad (error < 1,5\%) \tag{42a}$$

$$m_\tau c^2 = \frac{2m_e c^2}{\alpha}(1/\sqrt{\alpha} + 1) = 1779,5\,\mathrm{MeV} \quad (error < 0,2\%) \tag{42b}$$

giving a surprising agreement.

7 Masses of Heavy Vector Gauge Bosons

Heavy vector gauge bosons are the fundamental objects in the theory of electroweak interactions, consisting of the charged Woton W_+ and W_- , having the mass of $M_W c^2 = 80,419 \pm 0,056\,\mathrm{GeV}$ and the neutral Zoton Z_0 having the mass $M_Z c^2 = 91,1876 \pm 0,0021\,\mathrm{GeV}$.
If we want to extend the method of mass analogues to the case of the heavy vector gauge bosons, we have to look for vector particles, which could serve as reference particles. A promising choice seems to be the well known vector ρ-meson, having 3 charge states $\rho_\pm$ and ρ_0 , playing a fundamental role in the theory of electromagnetic form factors of the nucleon. But the problem is now, that ρ is a hadron, having in addition to the electroweak also strong interactions and that the idea of the mass analogues is so far only used for charged particles. To take this into account we introduce additional subtraction terms in the enhanced masses and make an ansatz in the form

$$m_{GB} = \frac{m_\rho}{\sqrt{\alpha}}\left(\frac{1}{\sqrt{\alpha}} - \beta(Z_e, Z_s)\right) \qquad m_\rho = 770\,\mathrm{MeV/c^2}, \tag{43}$$

where β is a quantity reflecting that an electric charge number Z_e and an analogous "strong charge number" Z_s will influence the mass subtraction. Discarding the strong interaction of ρ and extrapolating from the case of the heavy leptons we try for β the ansatz $\beta = 1 + |Z_e|$ giving the fictitious masses $\widetilde{m}_W = 87,4\,\mathrm{GeV/c^2}$ and $\widetilde{m}_Z = 96,4\,\mathrm{GeV/c^2}$, which are of course to high. But, surprisingly, using for the strong charge number the value $Z_s = 0,75$, which is of the same order like $|Z_e| = 1$, we get with $\beta = 1 + |Z_e| + |Z_s|$ the following results

$$m_W = \frac{m_\rho}{\sqrt{\alpha}}\left(\frac{1}{\sqrt{\alpha}} - (1 + 1 + 0,75)\right) \approx 80,6\,\mathrm{GeV/c^2}, \tag{44}$$

$$m_Z = \frac{m_\rho}{\sqrt{\alpha}}\left(\frac{1}{\sqrt{\alpha}} - (1 + 0,75)\right) = 89,6\,\mathrm{GeV}$$

which are in reasonable agreement with the observed values.
The terms subtracted from the enhanced mass due to $Z_s = 0,75$ are of the order of $8,77\,m_\rho$ respective $6,75\,\mathrm{GeV/c^2}$.

8 New Strategy by Redefinition of Sommerfelds Fine Structure Constant

With the help of the classical rest energy of the electron $E_{cl} = mc^2$ the binding energy of the electron in the hydrogen atom can be written $E_H = \frac{1}{2}mc^2\alpha^2$ providing $E_H/E_{cl} = \frac{1}{2}\alpha^2$. It is now advantageous to introduce a modified fine structure constant α_m defined by

$$\alpha_m^2 = E_H/E_{cl}\ ,\ \text{giving}\ \alpha_m = \alpha/\sqrt{2} = 5,16\,10^{-3}\ \text{and}\ 1/\alpha_m = 193,789 \approx .200 \tag{45}$$

It is now worthwhile to introduce in addition a modified electron mass m_e' by the relation

$$m_e'c^2 = 0,1\,\alpha_m\text{GeV} = 0,516\,\text{MeV}\ \text{deviating from}\ m_ec^2 = 0,511\,\text{MeV}\ \text{by}\ < 1\%. \tag{46}$$

Now we consider the strongly interacting charged pion with mass $m_\pi = 139,57\,\text{MeV/c}^2$, being related to the electron mass by the empirical relation

$$m_\pi \sim 2m_e/\alpha = 140,05\,\text{MeV/c}^2.$$

The corresponding modified pion mass m_π' is

$$m'_\pi = 2m_e'/\alpha = 2m_e'/\sqrt{2}\alpha_m = \sqrt{2}\cdot 0,1\frac{\alpha_m}{\alpha_m}\,\text{GeV/c}^2 \qquad \text{or}$$

$$m_\pi' = \frac{\sqrt{2}}{10}\,\text{GeV/c}^2 = 141,42\,\text{MeV/c}^2,$$

deviating from the correct mass of the pion by $< 1,3\%$.
But now $m_\pi' = \frac{\sqrt{2}}{10}\,\text{GeV/c}^2$ is independent of α_m, so no longer any connection exists to an electromagnetic coupling constant, meaning that at the mass of $0,141\,\text{GeV/c}^2$ a transition to another interaction, the strong interaction occurs! With the help of m_π' and the mass of the nucleon m_N (as input) a coupling constant can be determined, permitting to estimate the masses of the hadrons up to 1GeV/c^2 with a surprising accuracy.

References

[1] D. E. Groom et al, 2000, The European Physical Journal C15,1.

[2] C. Dekker, May 1999, Physics Today, 22 (further references are to be found there).

[3] G. Joos, 1989, Lehrbuch der theoretischen Physik, 15. Auflage, völlig neu bearbeitet von B. Fricke und K. Schäfer, AULA-Verlag, Wiesbaden.

[4] M. Niering, R. Holzworth, J. Reichert, P. Pokasow, Th. Udem, M. Weitz, T.W. Hänsch, P. Lemonde, G. Santarelli, M. Abgrall, P. Laurant, C. Salomon and A. Clairon, 2000, Phys. Rev. Lett. 84, 5496 (further references are to be found there).

Probabilistic Geometry and Information Content (an Introduction to Corob Theory)

Version 1.1 Jan 2003

Douglas J. Matzke, PhD
Lawrence Technologies, LLC, [1]
matzke@ieee.org

P. N. Lawrence, PhD [1]
nick@LT.com

Abstract

The topic of probabilistic geometry has been researched in the field of mathematics for well over 50 years. Applying the intrinsic yet unintuitive metrics in these high-dimensional spaces to the information arena is conceptually very tricky. Pentti Kanerva found results related to this field in the mid 80s and applied them to content addressable memories. Dr. P. Nick Lawrence also rediscovered a more generalized version of these results in the early 90s and has developed it into his patented computational theory called *Corobs*, which stands for *Cor*relational-Algorithm *Ob*jects. Recently, the link between quantum theory and Corob Theory was researched under DOD SBIR funding. This paper/presentation gives an overview of this field including the key concepts of how to implement useful computing knowing that randomly chosen points are all an expected or *standard distance* apart (robust equidistance of sqrt(N/6) in a real unit N-cube as N>20). These ideas are particularly relevant to the ANPA audience because of the direct application to "Program Universe" goals and results, since in a binary unit N-cube the distance and discrimination metrics are both XOR. For a link to the slides for this presentation see the Yahoo anpa-list email archive.

1. Introduction

The primary idea beneath ANPA's combinatorial hierarchy goals is the hope that known stable constants and laws of physics spontaneously emerge due to the combinatorics of low-level random information processes. As a result of years of research at Lawrence Technologies, LLC on random neurological processes and more recently random quantum processes, we have developed a high-dimensional, computation model we call "Corob Theory". This paper introduces the key ideas of Corob theory, which are related to the obscure mathematical field of probabilistic geometry. The goal of corob theory is to understand and mimic our ability to see and do things like previously seen and done. Corob theory and program universe goals are similar because both are interested in how something stable can emerge from purely random processes. Even though Corob theory started as a high-dimensional neurological based computation model, it was recently expanded to include the high-dimensional complex-valued

[1] 5485 Belt Line Rd, Suite 200, Dallas, Texas 75254, USA, see http://www.LT.com

quantum corobs. These results depend on the intrinsic and emergent yet unintuitive distance metric properties of high-dimensional spaces.

1.1 Approach and Concepts

The Corob Technology is based on the idea that neurology somehow uses randomness to do useful similarity computing, and understanding this approach will lead to computers that act more like living systems, thereby performing like synthetic organisms or Synthorgs™. This thinking led to the rediscovery of the little known mathematical property from probabilistic geometry [1] that randomly chosen points within a high-dimensional unit N-cube approaches a growing expected or standard Cartesian distance $\sqrt{N/6}$ as the number of dimensions grows large. Even though these results are statistical in nature, the standard deviation is a constant $\sqrt{7/120}$ = 0.2415, so effectively the actual range of the standard distance becomes extremely narrow and approaches the expected mean standard distance with high levels of certainty. We call this tendency toward the expected *standard distance* the fundamental Corob theorem.

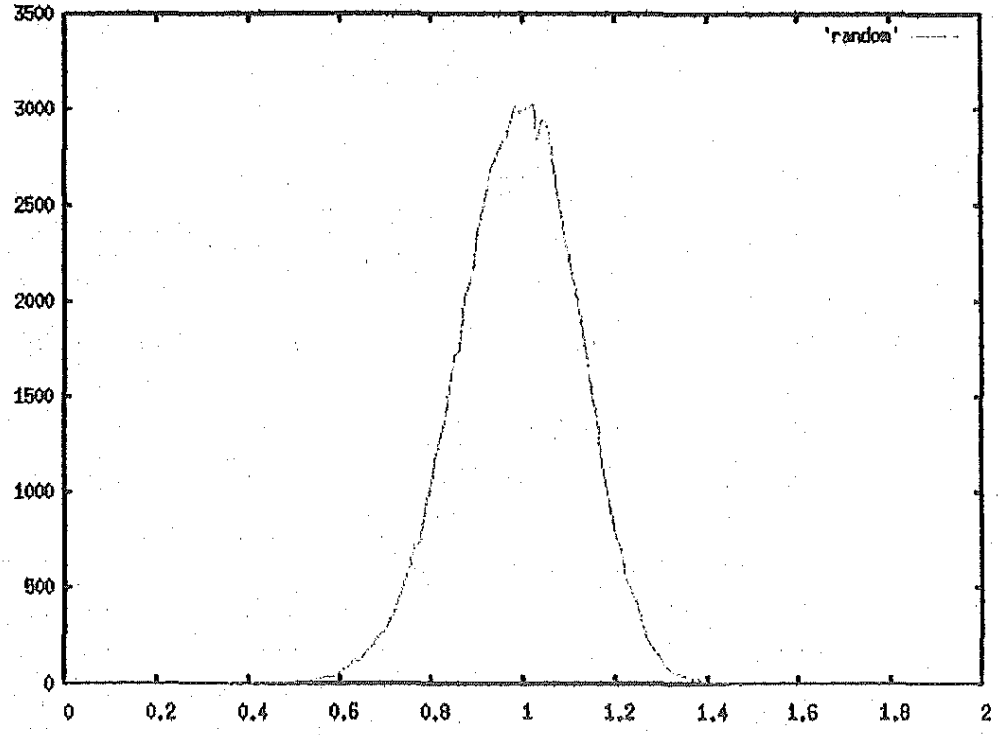


Figure 1. Random process with mean of 1.0

2. Randomness and Statistics

The results presented in this paper act like random processes with a mean and a standard deviation. Therefore, these two terms are introduced and discussed here. This section is a small refresher course on the key aspects of probability mathematics the corob theory relies on. Also, we define our own terms that rely on these statistical properties.

2.1 Mean or Expected Value

If a number of *uniformly distributed random values* are fed into *a random process* they produce a mean value with a high level of precision, so the random process tends to the mean, which represents the *expected value*. The random process actually generates values slightly smaller or larger than the mean, as shown in Figure 1. So even though the process tends to the expected value or mean, the actual values spread out around that center point. As one might expect random values close to the mean are more common than values further away. This spread can be characterized using the standard deviation and confidence interval, which are discussed next.

2.2 Standard Deviation or Dispersion

The exact amount of spread of the values from the random process is defined by the measure called *standard deviation*, which is the square root of the variance. Random processes naturally produce values closer to the mean value with a higher probability. The dispersion of values from the random process falls within the normalized interval defined by one *standard deviation* with a probability (or confidence interval) of 68.2%. This means that roughly two out of every three values falls within one standard deviation, nineteen out of every twenty values with within two standard deviations, etc (see Table 1). In contrast, it is possible but *extremely unlikely* that a random process will produce a value greater than 4 or 5 standard deviations away from the mean. This probability can be computed using the error function if the distribution is Gaussian. This relationship between probability and distance deviation from the mean has important information theoretic significance for the random processes discussed in this paper.

Table 1. Confidence Interval vs. standard deviation

Standard Deviations	Confidence Interval
±1	0.6826895
±2	0.9544997
±3	0.9973002
±4	0.9999366
±5	0.9999994

3. High Dimensional Spaces

The *primary random process* forming the basis of corob technology is simply applying the *Cartesian distance metric* between two sets of N uniformly distributed, bounded, random values, where each set represents the address of a random point inside a high-dimensional unit N-cube. This random process can be thought of in two ways, as a mean of the probabilistic process or an ensemble result similar to the concept of the center of gravity. The ensemble results and interpretation appears to be directly related to the probabilistic phenomena.

As this paper will demonstrate, the mean (and thereby the standard deviation) of several probabilistic Cartesian distances is related to the square root of the number of dimensions $\sim\sqrt{N}$, with some unexpected results as N grows much larger than the typical 3 Euclidean dimensions (N>>20). This chapter provides a step-by-step mathematical introduction to these main ideas.

3.1 Cartesian Distance Metric

The random process is simply the distance metric between any pair of points. Given an N dimensional unit N-cube and address values ($0 \le x_i \le 1$ and $0 \le y_i \le 1$) in each of the N dimensions for points $X = [x_1, x_2, \ldots, x_i, \ldots, x_N]$ and $Y = [y_1, y_2, \ldots, y_i, \ldots, y_N]$, then the Cartesian distance is simply the square root of the sum of differences squared between the two points or $d = \sqrt{\sum_i (x_i - y_i)^2} = \sqrt{\sum_i (z_i)^2}$. The distance to the origin in three dimensions (for X, Y and Z) is simply the well-known distance formula $d = \sqrt{(z_1)^2 + (z_2)^2 + (z_3)^2}$ or $d = \sqrt{(X)^2 + (Y)^2 + (Z)^2}$.

The primary unexpected result can now be discussed, which occurs when the values x_i and y_i are *randomly generated* with a uniform distribution random number generator, thereby creating a probabilistic geometry [1]. Under these conditions the distance between two randomly chosen points tends to an *expected constant standard distance* $\sim \sqrt{N/6}$ as the number of dimensions grows large N>>20. This equidistance tendency will be discussed next.

3.2 Tendency towards Equidistance

In three dimensions, the histogram of distances between randomly generated points X to the center point Y = [0.5, 0.5, ..., 0.5] looks like the graph in Figure 2. This histogram is basically triangular, where the asymmetry is due to fewer occurrences of the largest distances since they are non-uniformly located only in the extreme corners of the n-cube.

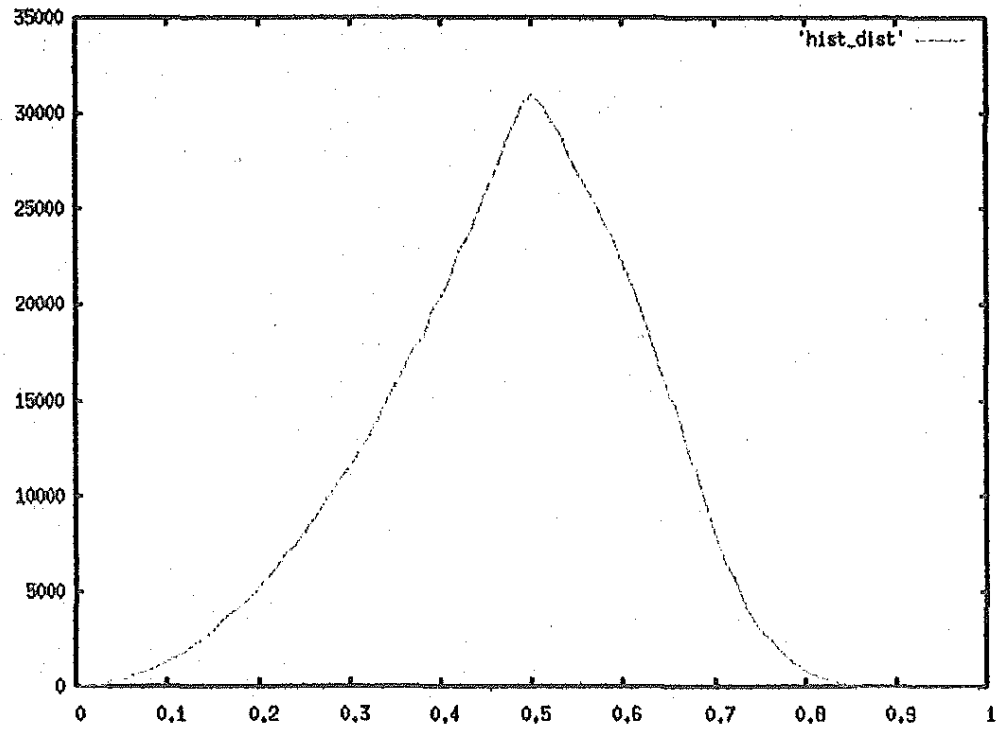


Figure 2. Distances for random points to center point in 3-space

The interesting thing to note is that Figure 2 is exactly the same random process as Figure 1 except that the number of dimensions representing each point is N = 3 rather than N = 12. The distances to the center point can be generalized for any number of dimensions and the result defines a probabilistic geometry, where any randomly chosen point probabilistically *tends* to be at the distance $\sim\sqrt{N/12}$ from the center of the space. This distance to the center is conceptually equivalent to an *expected standard radius* that grows with the $\sqrt{N}$ and the tendency toward that particular value grows very strong with increasing N because the *probabilistic distance metrics are dependent on* $\sqrt{N}$ but this *standard deviation is also a constant*!

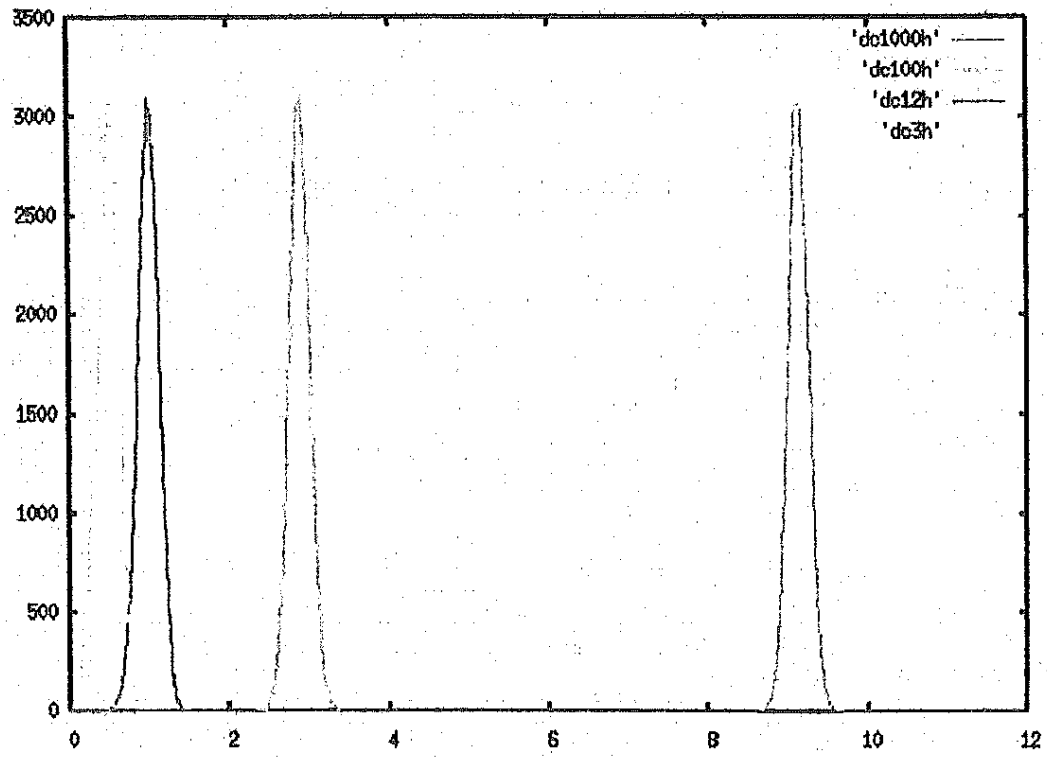


Figure 3. Histogram of Standard radius for N = 3, 12, 100 and 1000

Figure 3 illustrates the *standard radius* and dispersion of the values with increasing N. If the standard deviation is actually computed for each of the cases the standard deviation is the *constant value* $\sqrt{1/60}$. Since the standard deviation indicates the spread of values around the mean standard radius, the actual spread becomes a smaller and smaller percentage of the mean that grows with increasing N. Therefore, normalizing the standard deviation by dividing by the expected radius produces the ratio $\sqrt{1/60}/\sqrt{N/12} = \sqrt{1/60*12/N} = \sqrt{1/5N}$. The *normalized standard deviation actually shrinks* with N (and in the limit tends to 0 for very large N). This means the *standard radius tends strongly to a standard constant* value (for each N) with increasing N. See Figure 4, which is a normalized version of Figure 3. Because of this property, probabilistic geometries are easy to explore using simple programs with randomly chosen points, because they quickly converge to their respective mean values, and it is *high unlikely* to find a point that is 5-10 standard deviations away from the mean.

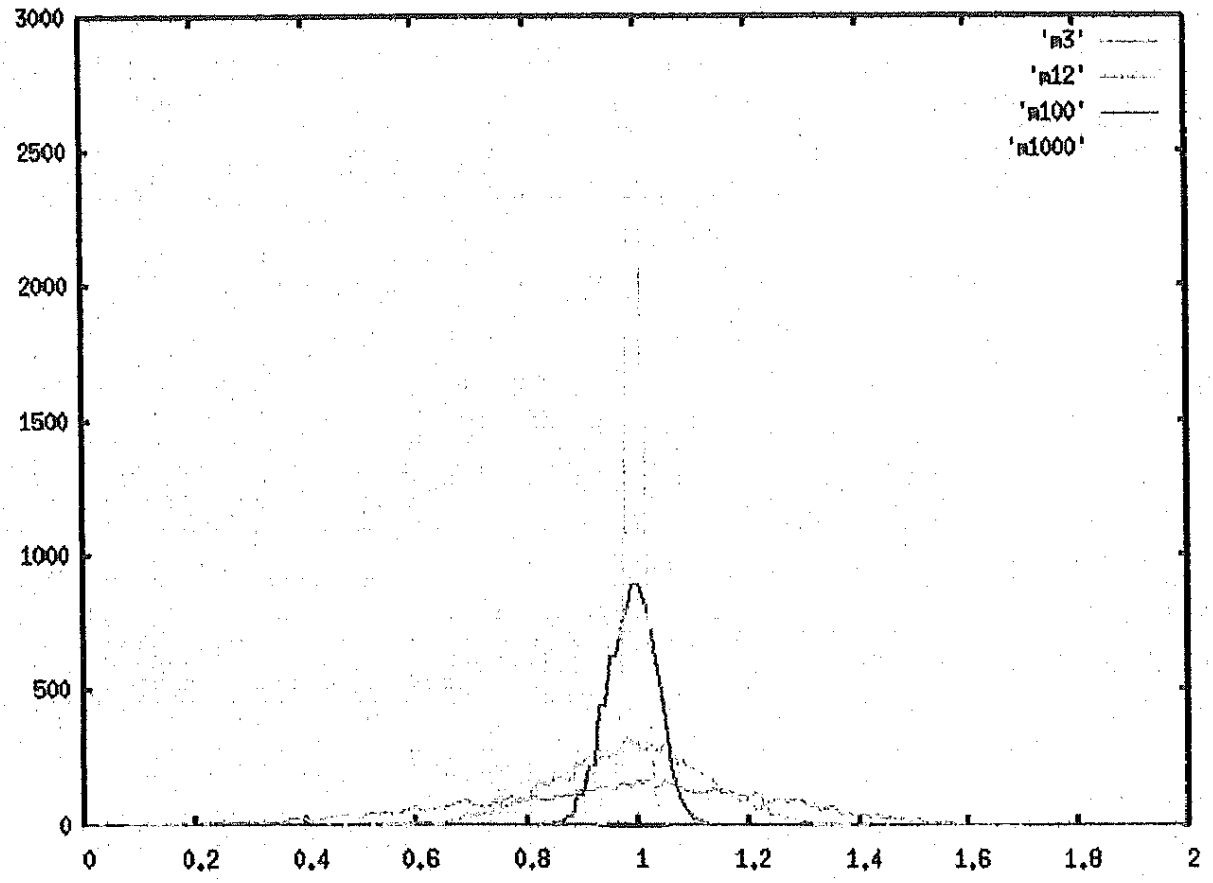


Figure 4. Normalized standard deviation of radius for N=3, 12, 100 and 1000

3.3 Probabilistic Geometries

The label of *standard radius* evokes the idea that the random points are located on the surface of a high-dimensional N-*sphere*. One way to attempt to confirm this is to find the distance between a 1000 randomly chosen points and the origin Y = [0, 0, …, 0] with N=1000 and the result is the distance $\sqrt{N/3}$, which happens to be twice the standard radius $2\sqrt{N/12} = \sqrt{4N/12} = \sqrt{N/3}$. This distance is unexpected because the known geometrical distance from the origin to the center is simply $d = \sqrt{\sum_i (0.5)^2} = \sqrt{N(0.25)} = \sqrt{N/4}$. Since the random point is actually the radius $\sqrt{N/12}$ away from the center, this defines a right triangle $\left(\sqrt{N/4}\right)^2 + \left(\sqrt{N/12}\right)^2 = \left(\sqrt{N/3}\right)^2$, which simplifies to N/4 + N/12 = 3N/12 + N/12= 4N /12 = N/3. Therefore the distance between any randomly chosen point to the origin is $\sqrt{N/3}$, but even more unexpected is that this is also the distance to *any random corner*!! Likewise, the standard deviation of this standard corner distance approaches 0 with increasing N. This result is hard to visualize because the distance between a random point and *every random corner* is the constant $\sqrt{N/3}$. This is similar to Pentti Kanerva's [2] distance between any two random corners of an n-cube $\sqrt{N/2}$ since the right

triangle from the center to two random corners has the property $\left(\sqrt{N/4}\right)^2+\left(\sqrt{N/4}\right)^2=\left(\sqrt{N/2}\right)^2$ and has the binomial distribution that can be derived using Pascal's triangle.

The most relevant statistical metric is the distance between *two randomly chosen points* inside the unit N-cube tends to the *standard distance* $\sqrt{N/6}$, with the standard deviation approaching 0 as N grows to infinity. It is not that the other points do not exist; it is they are *extremely* hard to find using random processes. For a large number of random points, they all tend to a *standard radius* from the center, a *standard diameter* (D=2r) from any corner and a *standard distance* from each other. Points with these properties *do not form* an N-sphere inside an N-cube, but rather a high-dimensional *tetrahedron* or N-*equihedron* (or N-shell), since all the points tend to be equidistant. This idea is illustrated in Figure 5, where all the blue points are equidistance from the center red dot (and more importantly from each other). If any point (big yellow dot) was then moved to the center, the big red dot would be among the remaining blue points. These distance properties are definitely wrong for an N-sphere so we use the term N-shell or N-*equihedron.*

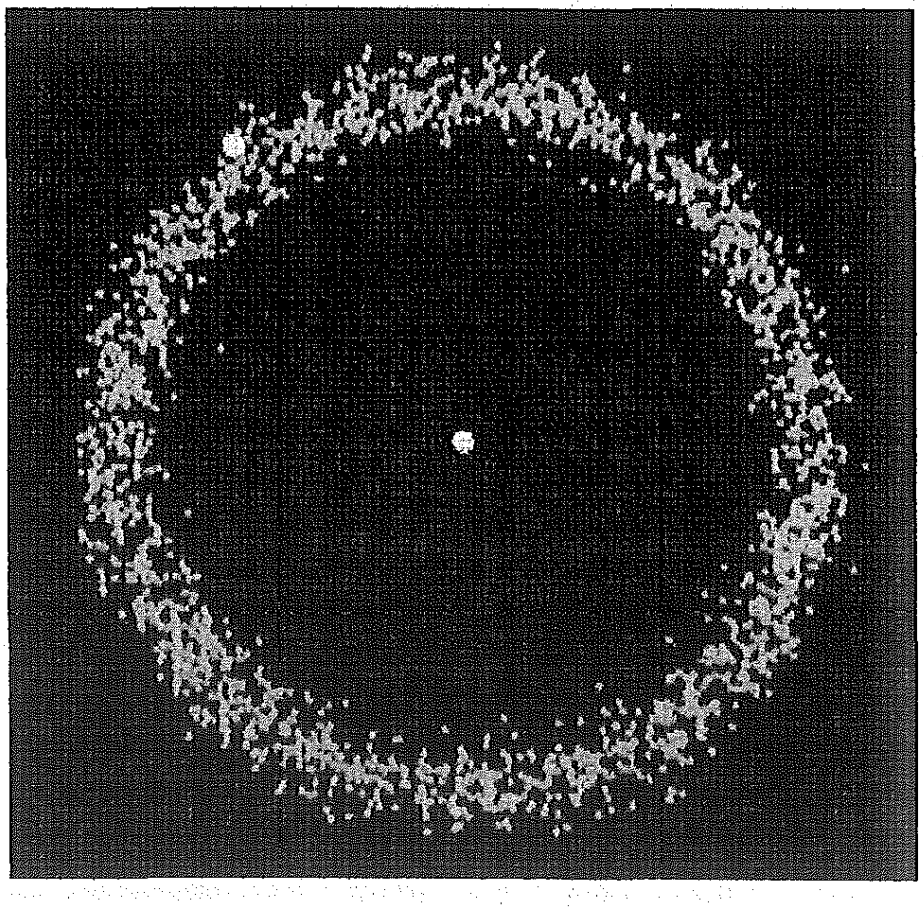


Figure 5. Planar Visualization of Equidistance Property

An alternative visualization of all the equidistance properties is shown in Figure 6, which depicts this topology using a planar equilateral triangle to represent the N-equihedron and all lines are distances normalized by the expected standard radius. Note that any lines of the same color are the same length even if it is impossible to draw them as such in a planar graphic.

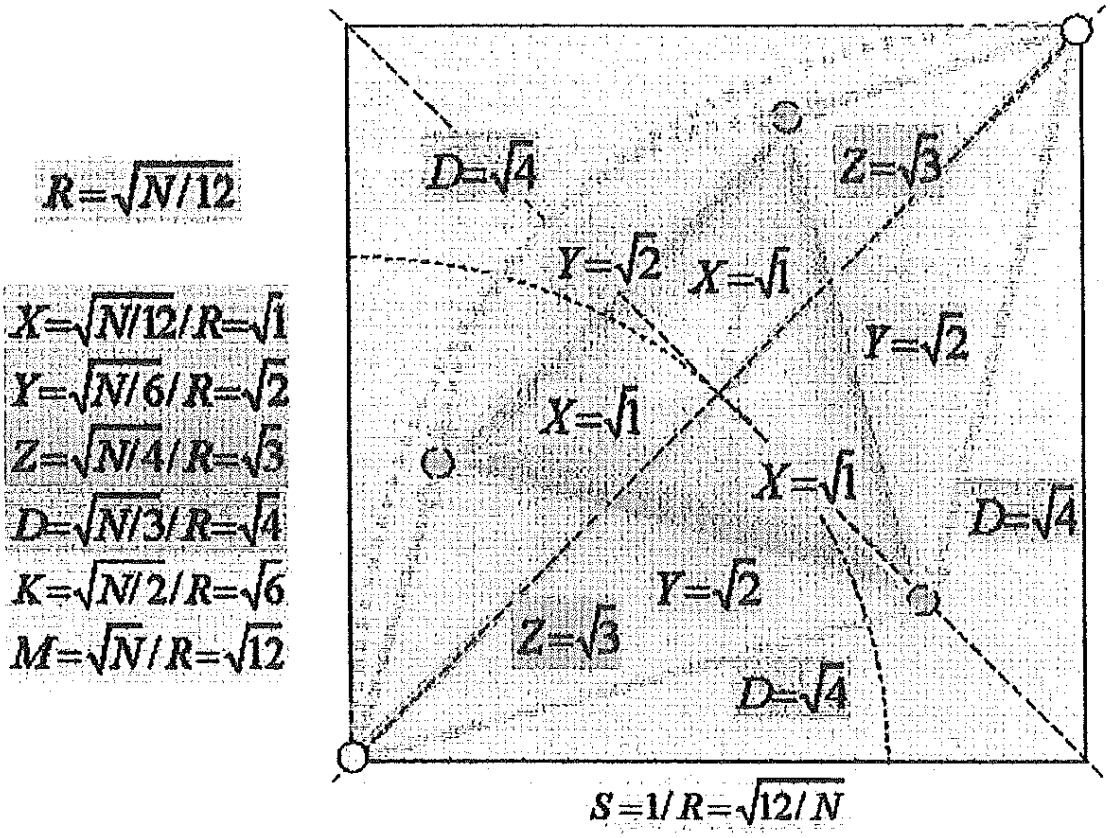


Figure 6. Standard distances from the perspective of green random points

An important fact regarding Figure 6 is all the distances have been normalized by dividing by the *expected standard radius* R resulting in *all standard distances becoming invariants or constants.* For example, with the normalized radius X = R/R ≈ 1, then similarly the expected standard distance is Y $\approx \sqrt{2}$, the normalized corner to center distance is Z $\approx \sqrt{3}$, the standard corner distance is D = 2R = 2 $\approx \sqrt{4}$ and the ncube's major diagonal is M $\approx \sqrt{12}$. The normalized Kanerva distance is K $\approx \sqrt{6}$. This means all expected distances are true *constants* independent of N and with vanishing standard deviations for very large N. Table 2 illustrates these same metrics in a tabular fashion depending on the preferred perspective of unit side or unit radius.

Since the standard radius and major diagonal are constants for any number of N, then the normalized length of every side S must shrink with increasing N or S = 1/R $\approx \sqrt{12/N}$. Maintaining a unit radius forces the N-cube to shrink and maintaining a unit side length on the N-cube means the standard radius increases proportional to $\sqrt{N}$. Since 12 is the magic number used for both the radius and Ncube side, when the number of dimensions N = 12 then the lengths R = S = 1 = volume. When the radius R = 1 then this acts like the unitarity constraint from quantum mechanics. This simple set of mathematical relationships is rich with potential meaning regarding the size or volume of things in both the classical and quantum domain depending on the number of dimensions and the normalization perspective. The number 12 is a special constant that emerges from this analysis, where these two aspects are indistinguishable.

Table 2. Geometric and probabilistic properties of unit N-cubes

Line	Type of Property	Distance with Unit Side	Distance with Unit Radius	Description of distance metric
1	Geom	$=1$	$\approx\sqrt{12/N}$	Side: any corner to a nearest corner
2	Prob	$R\approx\sqrt{N/12}$	≈ 1	Radius: midpoint to random point
3	Prob	$Y\approx\sqrt{N/6}$	$\approx\sqrt{2}$	Random point to random point
4	Geom	$Z=\sqrt{N/4}$	$\approx\sqrt{3}$	Any corner to midpoint
5	Prob	$D\approx\sqrt{N/3}$	$\approx\sqrt{4}$	Random corner to random point
6	Prob	$K\approx\sqrt{N/2}$	$\approx\sqrt{6}$	Random corner to random corner
7	Geom	$M=\sqrt{N}$	$\approx\sqrt{12}$	Any corner to most distant corner

A final point about Figure 6 is the arc drawn at distance $\sqrt{3}$ from the origin. Obviously there are more points outside this arc then inside this arc, so this represents why the ensemble length $\sqrt{3}$ is greater than the geometrical corner-center distance of $\sqrt{4}$.

This unit Ncube space is a distance metric space but nothing has been said about the angles. From standard geometry it is know that for right triangles then $a^2 + b^2 = h^2$ where sides a and b form a right angle and h is the hypotenuse. This property happens to be true for the expected standard distance $\sqrt{1}^2+\sqrt{1}^2=\sqrt{2}^2$ and the Kanerva distance $\sqrt{3}^2+\sqrt{3}^2=\sqrt{6}^2$, which suggests all randomly chosen points (or corners) are orthogonal to each other when treated as vectors from the center of the space. Obviously, the angles shown between the yellow lines of the standard radiuses all approach 90° even though Figure 6 cannot be drawn to scale. Since the standard distance is probabilistic, the angle between them is also not exact. The angle between to two vectors (defined as midpoint to each random point) approaches an *expected standard angle* of 90° as N grows. Figure 7 shows how the standard deviation shrinks with increasing N such that for N>3000 the standard deviation of the standard angle is <1°. The angle is computed with the formula in Figure 7 using the inner product $\langle x|y\rangle=\sqrt{\sum x_i y_i}$ and norm $\|x\|=\sqrt{\sum x_i^2}$ formulas.

$$\theta=\cos^{-1}\left(\frac{\langle x|y\rangle}{\|x\|\ \|y\|}\right)$$

Size of N	Standard Deviation	
	Inner Prod	As Angle
100	.1000	5.758°
1000	.0315	1.816°
10,000	.0100	0.563°

Figure 7. Expected Standard Angle approaches 90° for large N

3.4 Soft Tokens or Corobs

The entire probabilistic geometry is based on the premise that a set of random points generate these probabilities and at Lawrence Technologies we call such a random point a "corob" or soft token. A soft token is a valid term we have borrowed from communications theory, which has

the identical meaning, where the tokens are maximally spread out through an N-space providing an *error-correcting region* around each token. It is also know from coding theory the maximum information density occurs when the tokens are uniformly spread throughout the space and they are equally spaced. These properties naturally occur for corobs and have been known about in the field of probabilistic geometries for decades. The primary invention of corob theory is how to coax useful Turing complete computing from these randomly populated metric N-spaces.

3.5 Computing with Soft Tokens

Computing with random corobs requires defining tokens and assigning particular symbolic meaning to each of the tokens. This is equivalent to assigning *meaning* to a particular random address in the subspace of dimension N. All the other random points are extremely far way because the standard deviation of distance metrics are proportional to $\sqrt{1/N}$. Computation is simply mapping values from the environment into points in a subspace (using a sensor), finding which points are closest to *meaningful* known points, converting to another subspace (lobe), and finally converting the soft tokens back into the tradition numeric or string literals (actor). Corob computing is simply the process of creating spaces with the correct distance relationships to perform your desired behavior by moving the computation state around within the subspaces. This process can be shown to be Turing complete because lobes can define arbitrary logic gates and also represent dynamically stored state similar to a latch or cross-coupled NAND-gates. The following three sections discuss the various spatial and temporal relationship properties.

3.6 N-spaces and Subspaces

If an N-dimensional space is split in two, both of the resulting subspaces have the same properties regarding random tokens, as the whole N-space but with lower fidelity due to the increase in the normalized standard deviation. This is equivalent to the holographic property and can be easily understood as a noisy version of the original token. In general, random corobs are a set of random address values defined for a subspace of particular size N. Subspaces are concatenated together and sampled which effects tokens relative addresses depending on the desired operation. Remember that the directions up, down, right, left, east and west etc do not exist in subspaces. The only useful metric is similarity of distances compared to randomly chosen points, which were previously produced or known about in the system. For this reason, synthetic organisms respond in ways similar to previously seen, known or educated actions.

3.7 Cardinal, Ordinal and Continuous Relationships

Most of computing involves defining relationships between tokens, and corob theory geometrically represents all such relationships in N-spaces. The three kinds of relationships are:

1) Cardinal or *unordered* tokens (such as a list of names, properties or colors).
2) Ordered tokens (such as the alphabet where letter "c" is closer to "d" than "m")
3) Continuous spaces (such as 1-D for temperature and 2-D for images, etc)

Cardinal tokens are directly implemented using the random corobs already defined since they tend to be equally spaced with no apparent ordering. A subspace can hold M cardinal tokens

without any confusion by choosing the size N of the subspace such that some minimum spacing is guaranteed (i.e. 5-10 standard deviations) for that number of tokens.

Ordered tokens require that the distance between tokens is proportional to the difference in their adjacent sequence number. The distances between adjacent tokens must be smaller than standard distance and can be constructed using the following procedure. First decide what percentage of standard distance the tokens should be spaced apart (i.e. use s = 30% but depends on the application). Second, generate a random point $c_{i=i}$ and assign that to the first token "a". Third, generate another random point $c_{2=i+1}$ and use it as a random angle reference point at standard distance d. Interpolate between the two points to find a point close to c1 by using the formula $c_1*s + c_2(1-s)$ to compute the address of token "b". Repeat steps 2 & 3 between remaining points c_i and c_{i+1}. The resulting pearls on a string each have a distance metric that is smaller for its immediate token neighbor(s) than to the other tokens. When any token is compared to any of the pre-defined tokens, the closest tokens are easy to find. Also, the nearest neighbors are unique because it is extremely unlikely that random corobs will be less than the standard distance apart. This close spacing of sequential tokens is called a *string corob* (can also form a ring corob).

Continuous spaces are conceptually equivalent to ordered-tokens except that a continuous number of points must maintain the linear spacing as expected in a continuous variable. So if a continuous space is pre-populated with a set of ordered grid points, it is possible find an arbitrary number of interpolated points between each grid point. Unfortunately, the interpolated metric distances between grid points act as if they droop like fabric between tent poles because this process represents simulating a linear 1-D space within a high-dimensional Nsubspace with unusual probabilistic metrics. Therefore, a construct was invented and patented called the *toothpick method* for string corobs, which effectively randomizes the drooping effect. The string corob can be adapted for representing 1-D, 2-D, etc continuous spaces as illustrated in Figure 8.

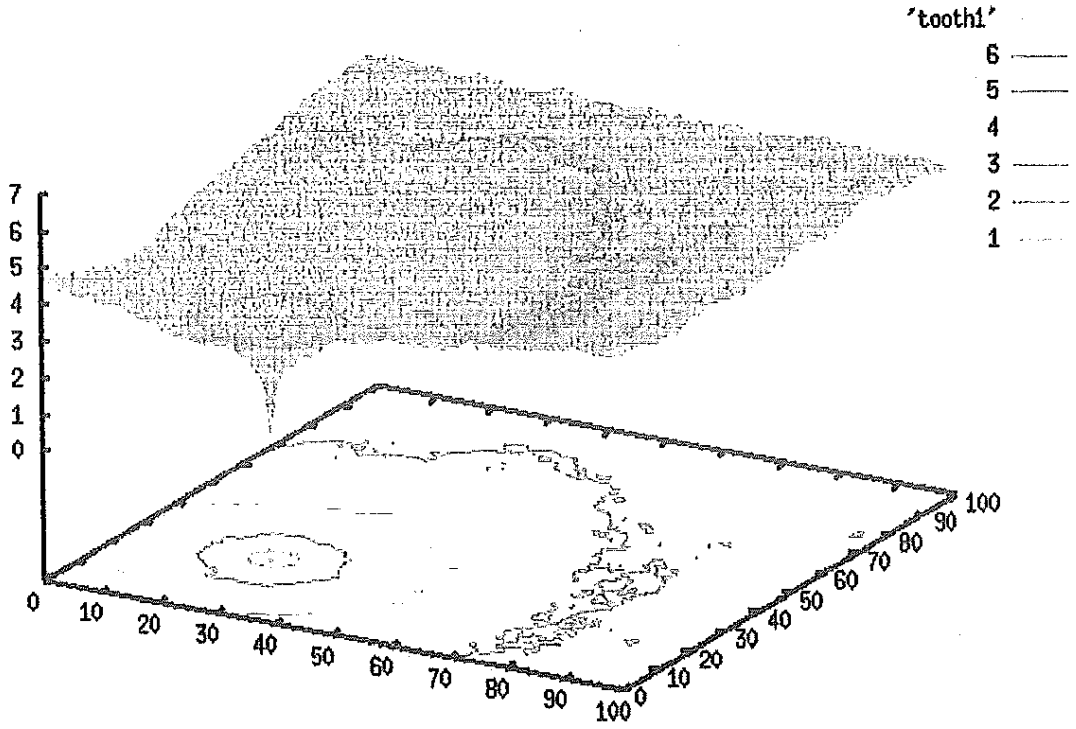


Figure 8. Embedding continuous geometries using string corobs and toothpicks

3.8 Mutual Exclusion and Concurrency

Corob tokens are stored in N-dimensional subspaces just like conventional classical tokens can be assigned a binary number and stored in a conventional n-bit register or memory. For example, the three tokens Red = 01_2, Green = 10_2 and Blue = 11_2 could be stored in a 2 bit register, which has the maximum capacity of $2^2 = 4$ possible distinct binary tokens. These distinct values are *mutually exclusive* because there is room for only one binary value at a time. Alternatively, the three tokens could be orthogonally coded as Red = 001_2, Green = 010_2 and Blue = 100_2 in a 3 bit register with a maximum capacity of $2^3 = 8$ distinct states, which then *simultaneously* allows any valid combination of Red, Green and Blue such as Purple = 101_2 and White = 111_2. Unfortunately, no other color shades can be represented in this encoding unless more intensity bits are allocated for each color (from 1 bit to 8 or 24 bits as used in modern computer graphics).

This discussion demonstrates that states (Red, Green or Blue) can share the same minimalistic bit encodings only if they are mutual exclusive. If multiple states must be simultaneously present then the encoding cannot be shared and must be redundant (orthogonal or separate bits allocated for each color). This same argument is true for tokens defined in subspaces in that each corob token is encoded as a random address in the subspace. The difference between subspaces and registers is subspaces naturally have extra room to simultaneously represent mixtures of naturally orthonormal tokens using the principles of vector addition.

For example, if Red, Green and Blue are each assigned a random corob in the same subspace (size N = 20 gives statistically reasonable separation) then they can be thought of as orthonormal states with an error correction region around each point. It is also possible to create points that are partway between Red and Blue (just average their addresses) and form a color we know as Purple. So even though the colors Purple and Yellow etc were not predefined for the "color subspace" they can be represented in the system as a concurrent mixture of other primary tokens.

This kind of concurrency or token mixtures does not emerge in the traditional mutual exclusive binary encodings of tokens unless they are separately coded. Because of the minimal encoding, conventional binary encoding have abrupt edges while corobs represent soft tokens with a unique kind of concurrency encoding possible. This concurrency can occur because corob points can also be thought of as orthonormal basis vectors. This concurrency is also similar to the mixture of spin states in quantum computing, thereby leading the qubit property we call superposition. The primary idea behind superposition is that the spin states are NOT mutually exclusive, so the system state can simultaneously represent both Up and Down or both True and False.

3.9 Noise Insensitivity and Reliability

Most tokens are assigned a random address in a subspace, which are naturally far apart. If some process creates a noisy version of one of the tokens (drifting away in a random direction by mixing in another unknown random token) the distance metric will still identify the closest original token compared to all the other known tokens up to the standard distance limit. This probabilistic geometry property leads to reliability of identifying tokens in the face of noise less than the amount equivalent to pure randomness. This idea that the amount of noise is directly quantifiable based on the standard distance metric is an important idea central to corob theory.

4. Distance and Information Content

4.1 Distance Metric is Information

Any probabilistic process defines a mean and standard deviation. In addition, the confidence interval is proportional to the standard deviation and can be computed using the error function (by assuming the distribution as a Gaussian). This means that a distance metric represents a probability, which can be directly converted to the equivalent information content in bits. Therefore two points at the standard distance have little information content since they are at the expected distance. Alternatively, two points that are close compared to standard distance represent a much higher information content since it is unexpected due to random processes.

4.2 Information content of each dimension

The information content of each dimension can be computed by starting with the standard distance represented as the number of standard deviations $\sqrt{N/6}/\sqrt{7/120} = \sqrt{20N/7}$, which leads to a probability using the error function prob = erfcc($\sqrt{20N/7}$. The final bit content of this standard distance probability is: bits = $-\ln_2$(prob). Figure 9 shows the plot of these equations for N from 1 through 250, where after that the error function (which is an integral) underflows for double precision floating-point numbers.

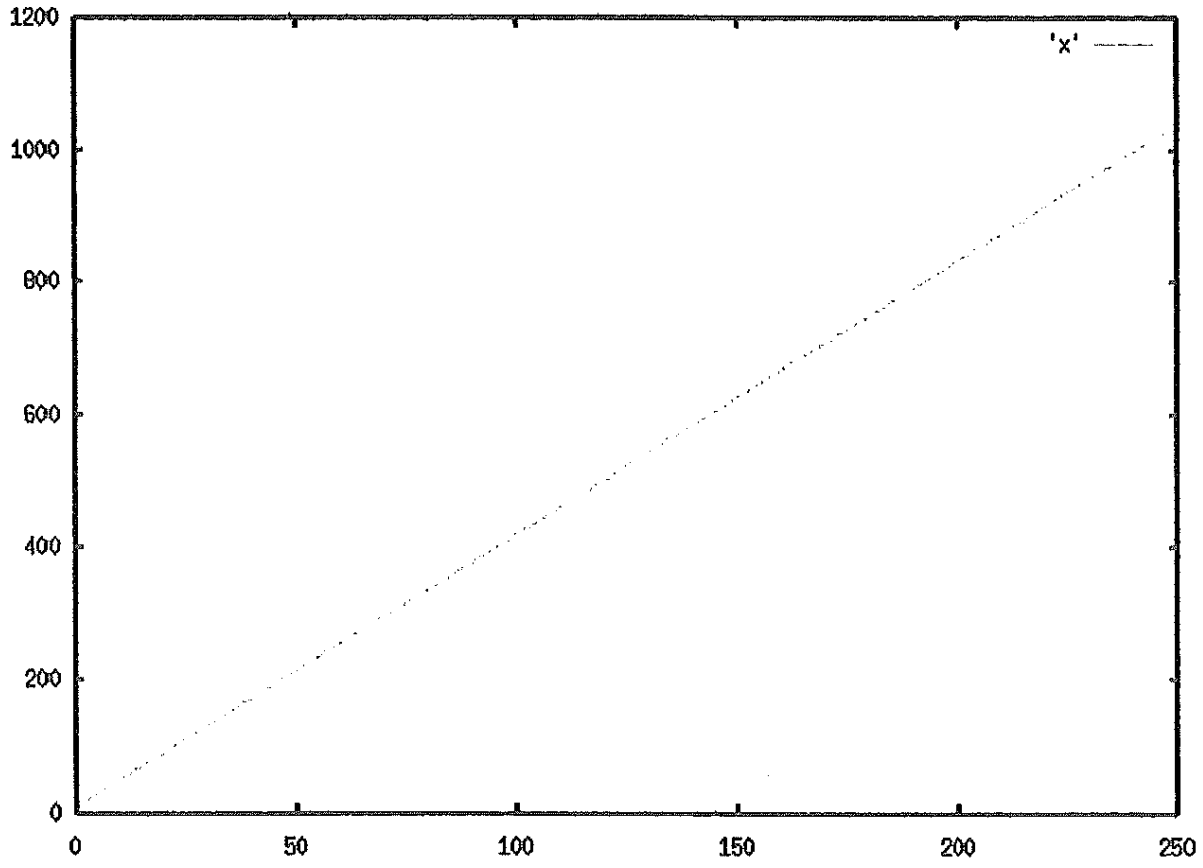


Figure 9. Bit content of standard distance for N=1-250

This plot produces a straight line that approximates the equation of $1+\left(N/\sqrt{7/120}\right)=$ $=1+(4.14*N)$. So each dimension (a floating point number $0 \le x_i \le 1$) is worth slightly more than 4 bits of information or roughly sixteen discrete values in each dimension. This makes sense since the N-space volume and the mean standard distance grows with N, so more bits are required to represent probability in larger spaces since they are effectively N-shell ratios. This limit to the precision needed for representing each dimension is primarily due to the distance metric, which sums many such squared dimensions resulting in the lower order bits becoming insignificant. This also means that the hardware accelerators for this kind of system may be able to use less than 6-8 bits of precision for each adder or multiplier used in the squaring operations.

Based on Figure 9 and the equation $1+N\sqrt{120/7}$ (or approximately $1+N(1+\pi)$) it is possible to represent either probabilities or bit content by creating specifically sized spaces to hold that information content. Pre-specified grid points can be carefully located in that space and then a test point can be created with multiple *simultaneous probabilities*. This strong relationship between *concurrent probabilities*, information content and distance metric is expected to be useful in the quantum-computing arena.

5. Quantum Corobs

The key step in applying corob theory to quantum computing is to compute the standard metrics for bounded, complex-valued spaces. The first step is to make the space symmetric around zero by choosing the bounds to be −1 to +1 thereby making the side length of S=2 and the space center the point [0,0,0,...]. All distance metrics normalized by the standard radius are the same due to axis relabeling. The next step converts real to complex numbers by transforming uniformly distributed random values to uniformly distributed random phases. Finally, we must deal with the fact that the vector dements are complex numbers. So the difference between complex numbers produces $z_i=(x_i-y_i)$, which can be used in Cartesian distance formula of $\sqrt{z_1^2+z_2^2+...+z_N^2}$. This requires the complex square of $|z_i|^2=z_i^* z_i$, where $z_i^*=a_i-ib_i$ is the complex conjugate of $z_i=a_i+ib_i$ resulting in $|z_i|^2=(a_i-ib_i)(a_i+ib_i)=a_i^2+b_i^2$, which is the scalar magnitude squared of the complex number.

When a large array of N uniformly distributed phase complex numbers with magnitude = 1 is placed through discovery of the standard metrics, the results are shown in Figure 10. This array of N complex numbers also exhibits the corob standard distance metrics. Notice how the corner distance distribution is different due to the constraint that each complex magnitude = 1. In any case the standard distance is the same after renormalization using the standard radius. Arrays of complex numbers form the essential mathematics of quantum Hilbert spaces.

Both the real and complex spaces that are symmetric around point [0,0,0,...] have the property that randomly chosen points also naturally represent a vector from the center, without any fix up. This is useful since the same representation can be used for distance metrics or vector computations (i.e. inner product). Also the center does not shift if the bounds of the space grows.

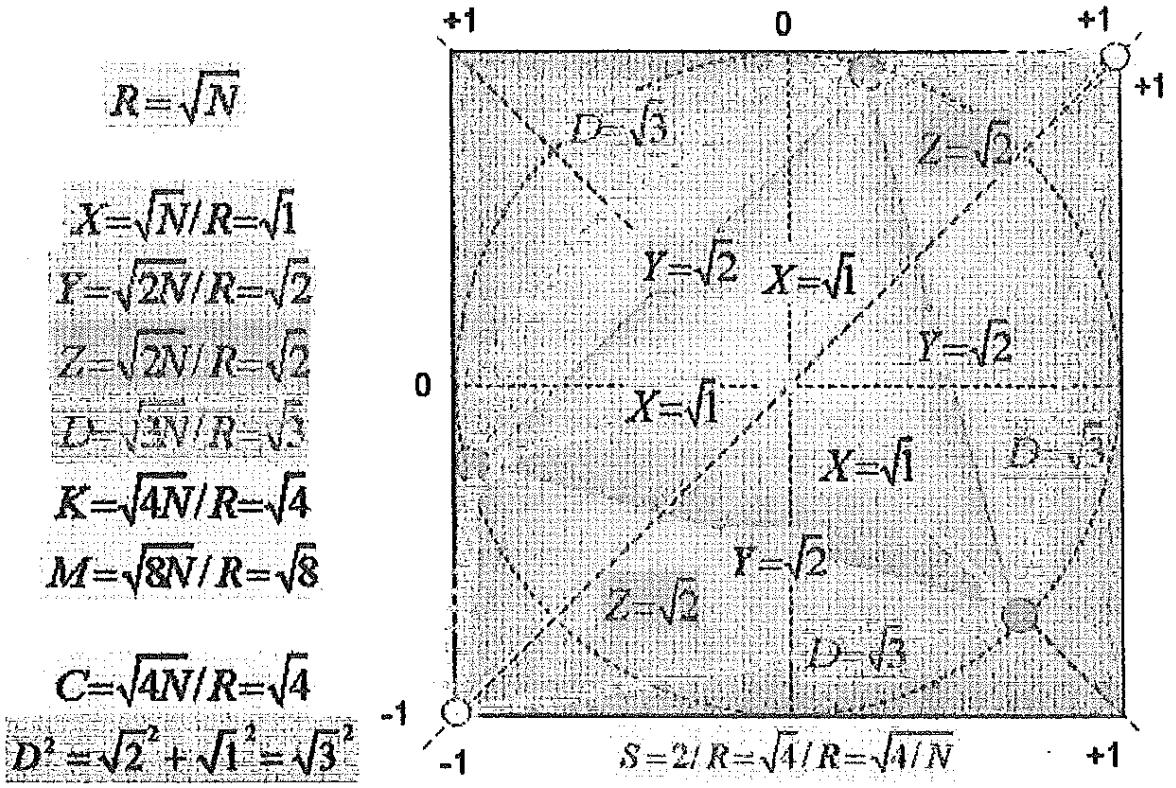


Figure 10. Uniformly distributed phase and magnitude = 1 complex numbers for N_c

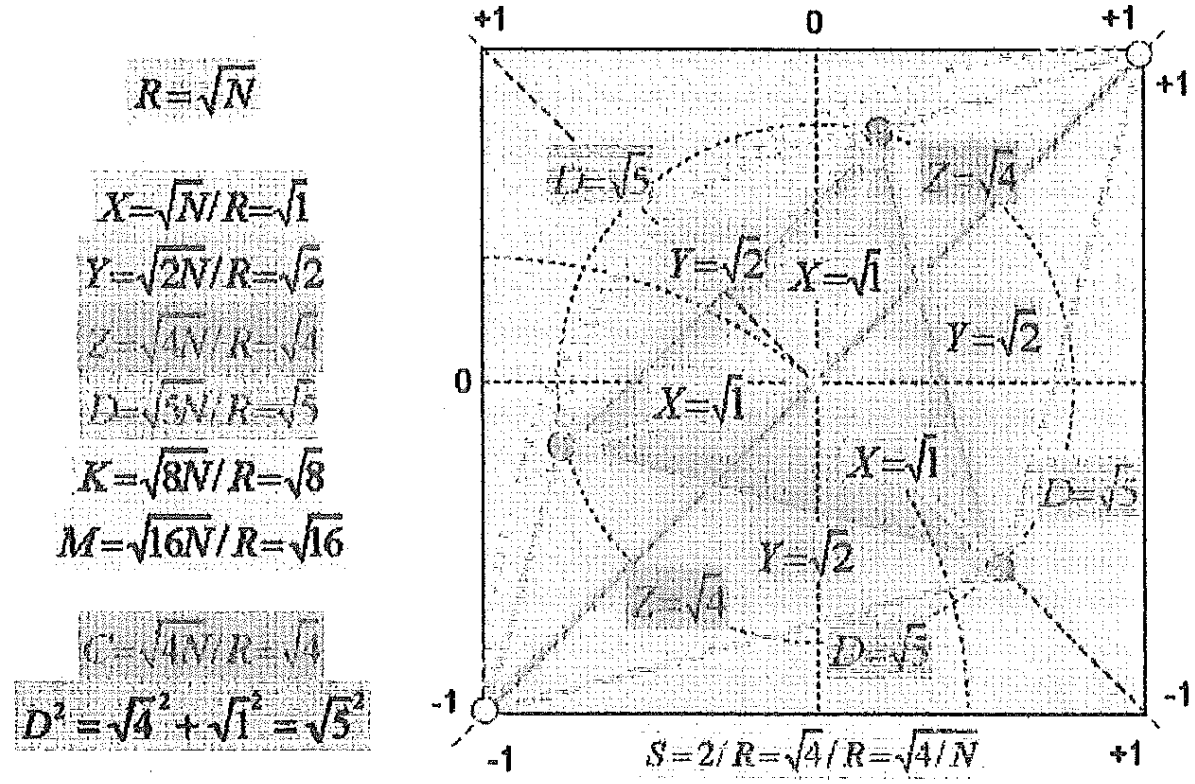


Figure 11. Uniformly distributed phase qubits for N_q or ebits for N_e with q=2

This same standard metrics discovery process can also be applied to arrays of mathematical objects of higher grade, such as isolated arrays of qubits and arrays of ebits. Figure 11 shows the same normalized distance metrics that result for both. These arrays of qubits are not the typical entangled quantum registers but simple arrays of *isolated qubits*, which we call a quantum ensemble. The arrays of ebits are also arrays of isolated ebits, where each ebit is formed by inseparably entangling two qubits. Quantum corobs are quantum ensembles formed on top of simple arrays of qubits or ebits, where the phase is the primary random variable.

5.1 Measurement as Noise Injection

It is possible to encode classical corobs into quantum corobs by converting random values to random phases. It is also possible to convert a quantum corob back into a classical corob or Kanerva bit array by *measuring* the quantum corob. The key result of this measurement process is the quantum corob survives this process. We believe this result is so significant that we have already filed a provisional US patent on the idea. Here are the details.

The key concept about this measurement process is when evenly distributed phase angles are projected; they produce results that look like a noisy version of some average result. Most importantly, the noise is smaller than the amount that would make it appear like another *completely* random point. In simple corob terms, the distance between repeated projections of a repeatable corob soft token X is smaller than standard distance, whereas the distance between projection of soft token X compared to that of soft token Y is at appropriate standard distance. Figure 12 pictorially represents these relative distances of repeatable projections.

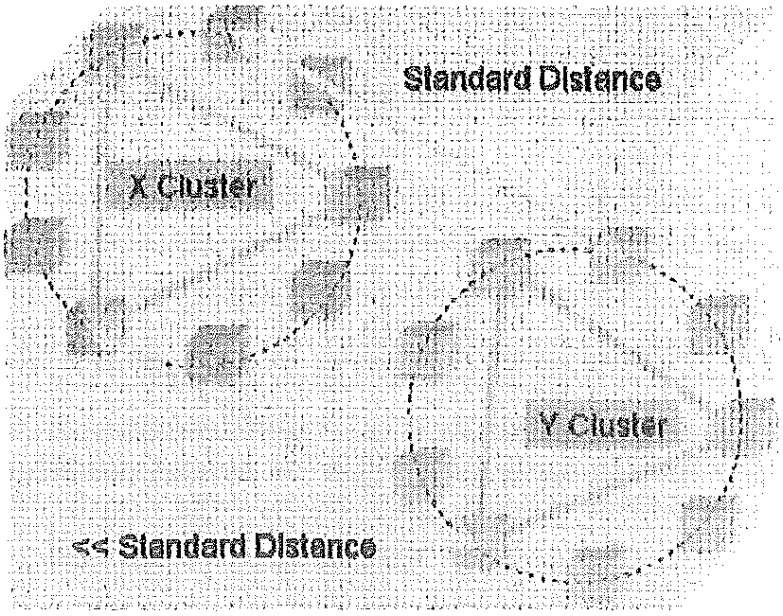


Figure 12. Cluster Projection distance for qubit arrays

Figure 13 shows the histogram that proves the earlier statement that repeated projections from a repeatable qubit soft token X acts like noisy version of some average expectation result because the distance is roughly 0.5 of standard distance. The rule of thumb we use is if the standard distance is 0.5 then the number of non-noisy dimensions is $(0.5)^2 = 0.25$ to give the same

distance, so roughly 25% of the dimensions are noisy. This is possible even with arrays of random phase qubits because the projection process favors points/probabilities that are closer to either measurement basis set compared to the purely random results from phases 45 degrees away. Since the initial phase is totally random the choice of basis states is also immaterial. The histogram demonstrates that the standard deviation also shrinks as N becomes larger. The labels mean; ic = "in cluster" and bc = "between cluster". These histograms indicate it is highly unlikely for expected values around 0.5 standard distance will ever be confused with points at standard distance, therefore repeatable, qubit-encoded quantum corobs survive measurement.

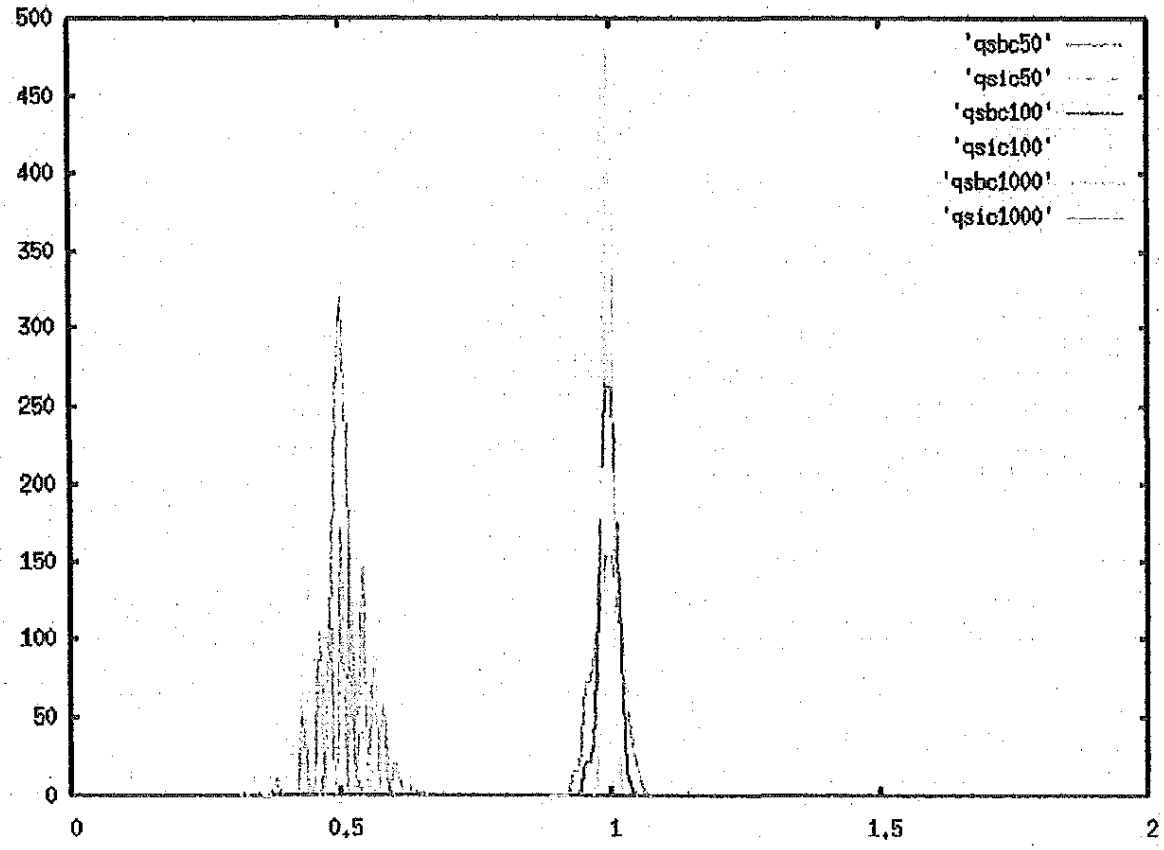


Figure 13. State cluster measurement histogram for qubit arrays N=50, 100, 1000

This result is interesting and quite unintuitive, because quantum randomness is not considered to be a computational resource. The key to this result is that isolated quantum ensembles must be used instead of entangled quantum registers. Notice discrete states when N=50 is low.

This result can be easily validated by looking at arrays of larger grain-size mathematical objects vs. complex numbers, qubits or ebits. Assume that each array element is an entangled quantum register of Q=2, 3, 4, or more qubits. Using the tensor product to create entangled qubits forms larger sets of unique basis vectors. Table 3 shows the result of taking a measurement using an arbitrary phase basis vector. The key conclusion is that the larger register the more random the measurement process produces. Even though quantum registers are producing corobs, they *generate random corobs* rather than tokens that survive. These numbers are slightly different than the talk slides due to improvements in our libraries of generators.

Table 3. State Cluster measurement distances for arrays of higher grained qubit registers

Element	% of standist	# of reals	Entangled state equation
complex	70.7 %	2	a (is complex)
1 qubit	50 %	4	$a\|0\rangle+b\|1\rangle$
2 qubits	65 %	8	$a\|00_2\rangle+b\|01_2\rangle+c\|10_2\rangle+d\|11_2\rangle$
3 qubits	76 %	12	$c_0\|000_2\rangle+c_1\|001_2\rangle+...+c_6\|110_2\rangle+c_7\|111_2\rangle$
4 qubits	83 %	16	$c_0\|0000_2\rangle+c_1\|0001_2\rangle+...+c_{14}\|1110_2\rangle+c_{15}\|1111_2\rangle$
>4qubits	100 %	>16	$c_0\|0\rangle+c_1\|1\rangle+...+c_{N-2}\|N-2\rangle+c_{N-1}\|N-1\rangle$

This same result is illustrated graphically in Figure 14 for a fixed N=1000 with changing grain size of q=1 vs. q=2. These measurements are the cluster distances of the new state results after measurement.

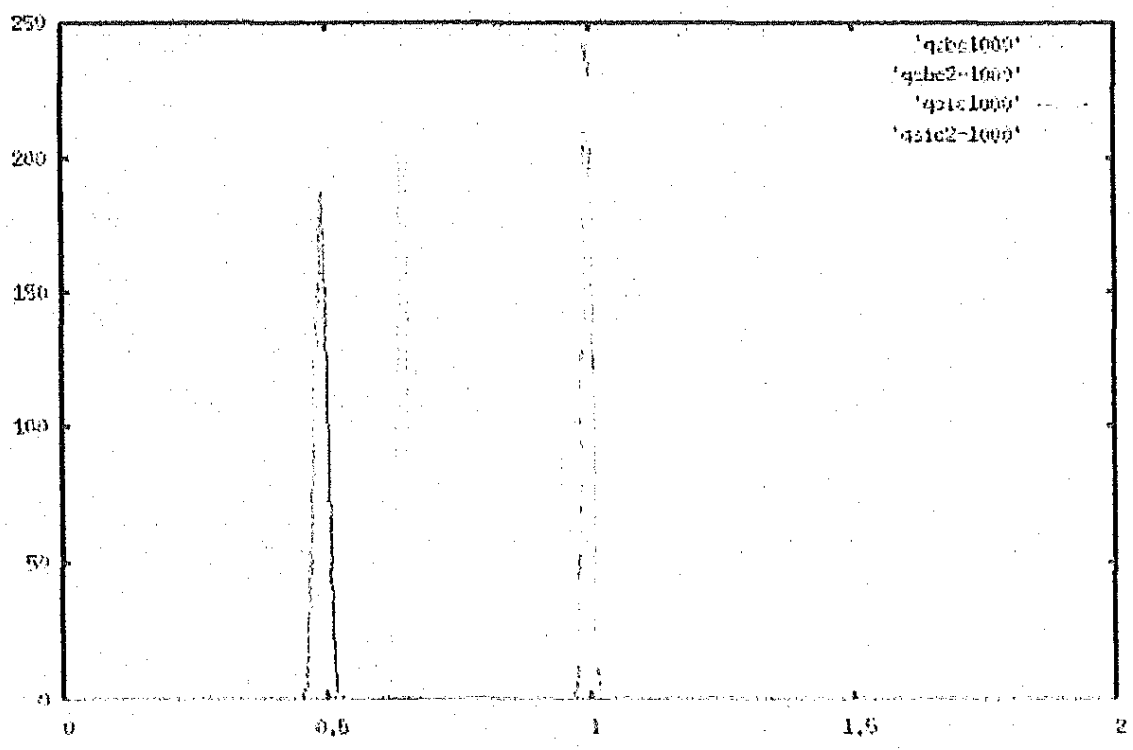


Figure 14. State Cluster histograms for arrays N=1000 of qubit registers with q=1 and q=2

The new resulting state is only part of the measurement process. The *answer* can also be viewed as a corob state since it is a binary valued Kanerva point. If many answers are averaged together they form the expected value, which is just the probability of finding a 1, so can be represented as real-valued corob. Figure 15 shows the histogram distributions of the measurement answers.

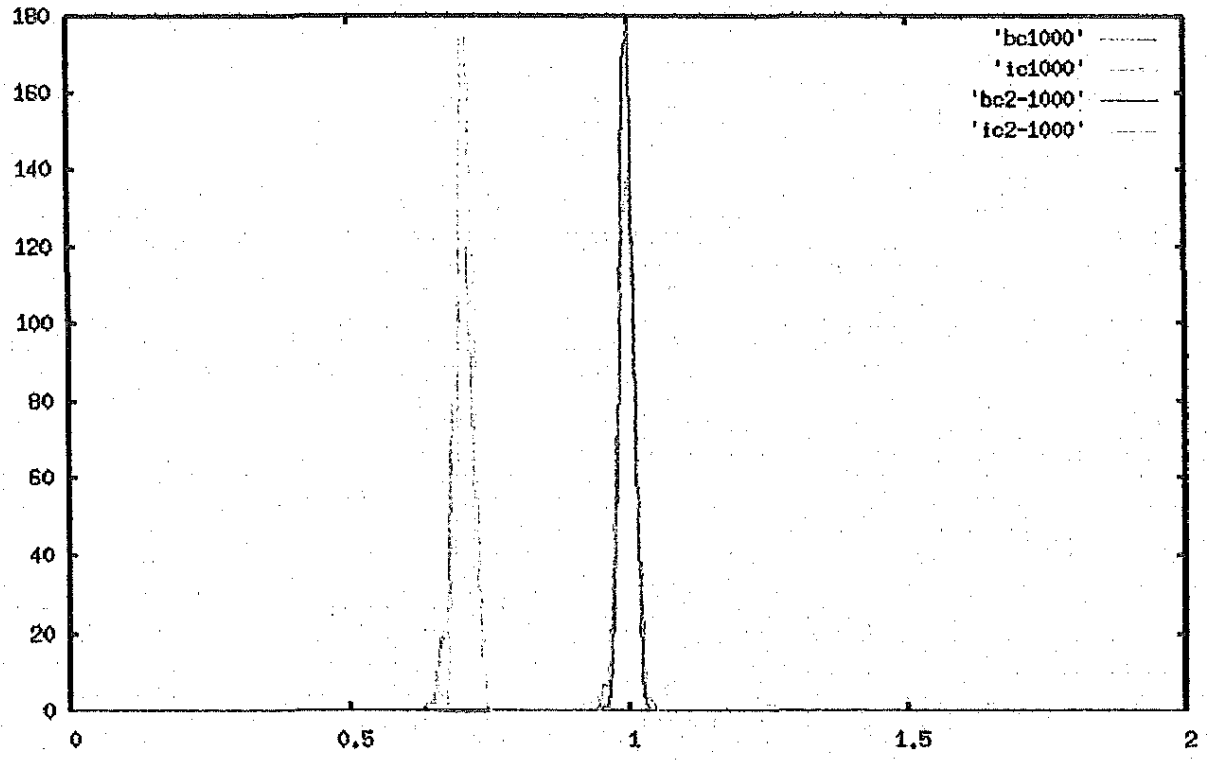


Figure 15. Answer cluster histograms for arrays N=500 of qubit registers with q=1 and q=2

In conclusion, isolated arrays of qubits (or quantum ensembles) allow representing quantum corobs that survive measurement, whereas quantum registers produce random corobs.

5.2 Ebit corobs survive measurement

Constructing arrays of uniformly distributed random phase complex numbers or qubits (pairs of complex numbers) is mathematically straightforward. A qubit is represented as expression $a|0\rangle+b|1\rangle$ where complex numbers a and b must meet the unitarity constraint of $a^2+b^2=1$. The first step in investigating corobs using ebits is that generating uniformly distributed arrays of random phase ebits is not as simple. These ebit states are the *inseparable entanglement* of two individual qubits. With complex numbers a, b, c and d an ebit has the representation of $a|00\rangle+d|11\rangle$ (or $b|01\rangle+c|10\rangle$) with respective unitarity constraints of $a^2+d^2=1$ with $b=c=0$ (or $b^2+c^2=1$ with $a=d=0$), which both meet the overall constraint of $a^2+b^2+c^2+d^2=1$. Notice how no tensor product of two smaller states can produce either combination of states, which is the definition of inseparability for ebits. It is still possible to mathematically produce a uniform random phase distribution of ebits after the tensor product and erasure steps.

When an ebit is measured, the two qubits are then broken apart into two separate unentangled qubits of the appropriate states. So the measurement for an ebit array of size N produces two qubit arrays of size N and a binary answer array of size N. These can all be analyzed from a standard distance metrics with the following results. Again Figures 16 and 17 are slightly different than the presented slides due to improvements in the ebit modeling libraries.

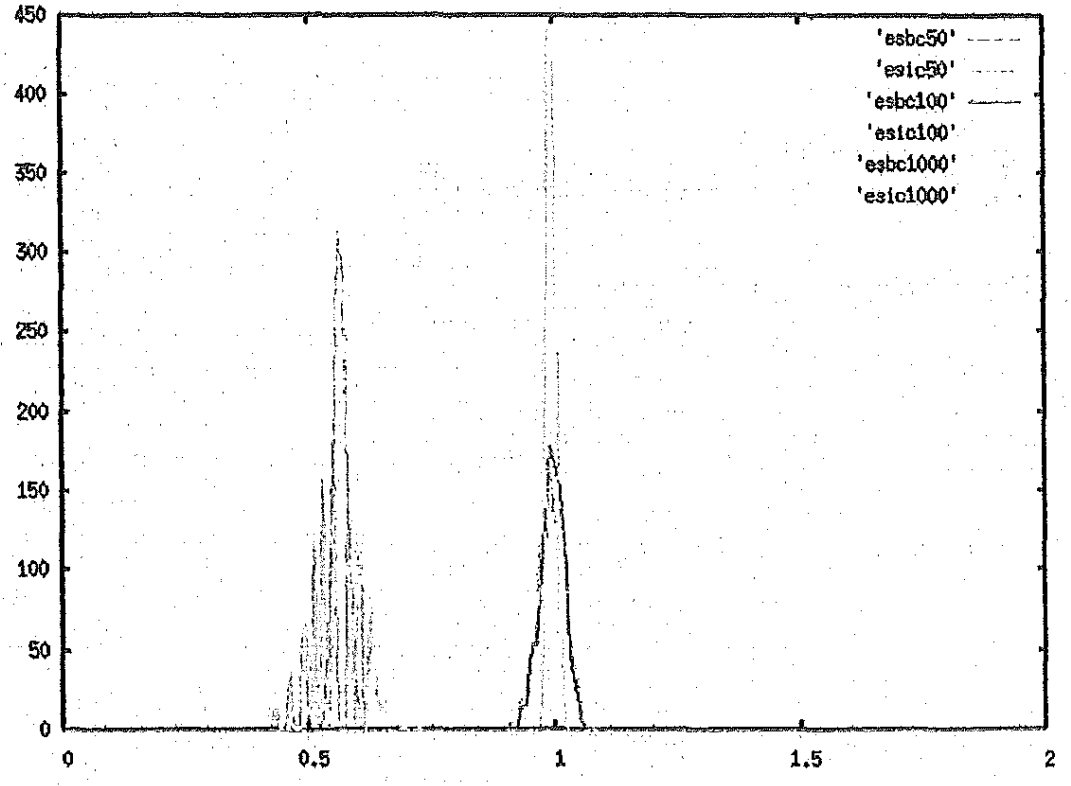


Figure 16. State cluster distance histogram for ebit arrays (q=2)

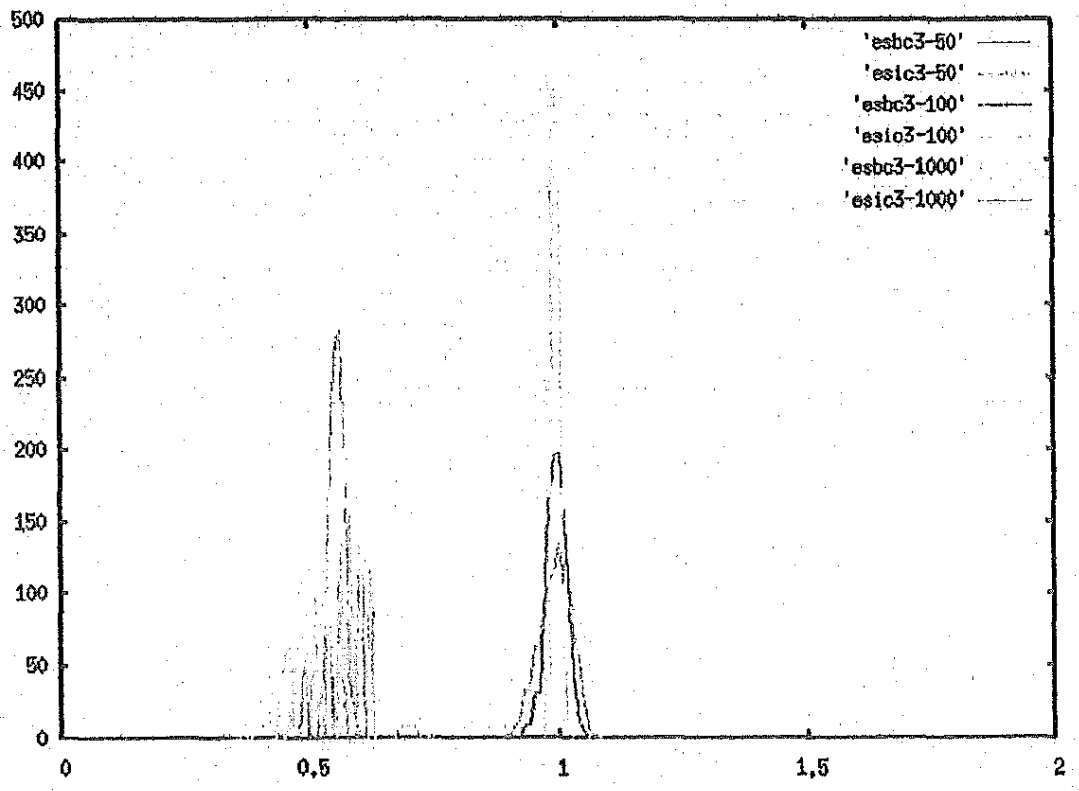


Figure 17. State cluster distance histogram for ebit arrays (q=3)

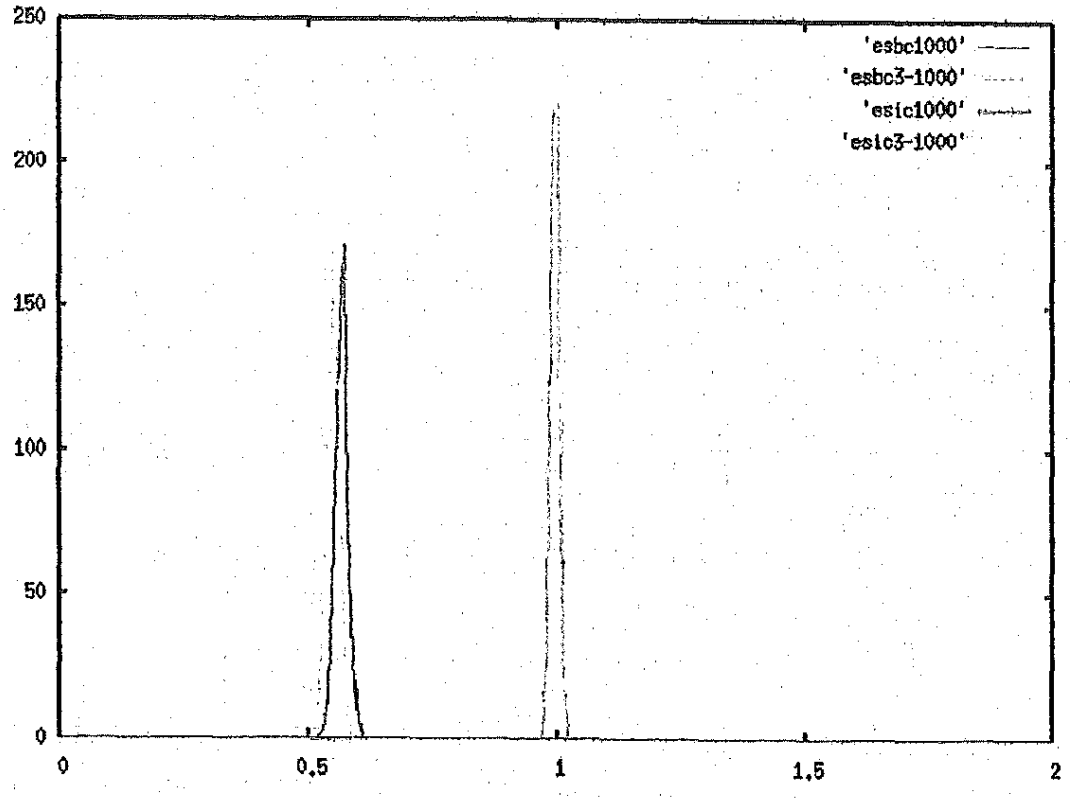


Figure 18. State cluster distances for ebit arrays N=1000 with q=2 and q=3

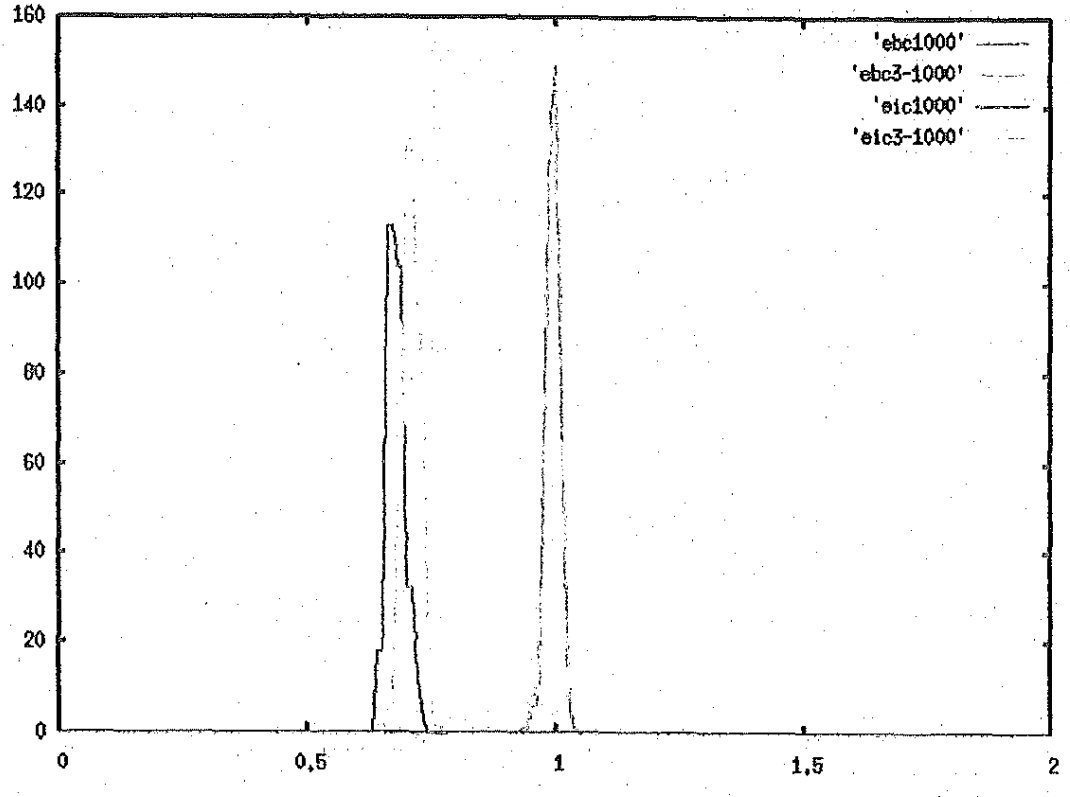


Figure 19. Answer cluster distances for ebit arrays N=1000 with q=2 and q=3

The overall result of measurement of quantum corobs represented in ebit arrays, is the both the answer and the qubit results indicate survival of the tokens. Quantum corobs survive measurement. Measurement of highly entangled quantum registers produce random corobs.

6. Conclusions

The striking result of this research is that noise and quantum randomness can be quantitatively viewed from the corob perspective. This result shows that unentangled quantum ensembles can be useful for representing unique data tokens in both the classical and quantum domains. Most importantly, these quantum data tokens can survive quantum measurement, which may be useful for any neurological theory that would make use of quantum mechanisms.

In order to explore this interrelation between classical and quantum corobs, we are building a corob language and tool suite that would allow systems to be defined containing arrays of both types of cells. This language is currently going thru the second design revision. The key computational mechanism is lightweight concurrent rules that act as unique observer perspectives defined by corobs or basis states in various subspaces. These rules appear to have the characteristics of quantum superposition because states can evolve in multiple simultaneous paths. It is expected that corob language with quantum extensions would allow research in interesting quantum domains such as quantum key distribution and quantum computation.

These results are novel and interesting since most people researching quantum computing do not know about corob theory. Anytime a large number of independent random values are represented by a system then standard distance metrics emerge. Program universe also has these same properties so therefore corob distance metrics must naturally emerge from that endeavor. The corob theory also is quite instructive about how stability can emerge from purely random processes (since the discrimination metric is identical to distance metric), which is virtually the same goal as Program Universe.

References

1. D. Klain, G. Rota, Introduction to Geometric Probability, Cambridge University Press, 1999. Also see http://pauillac.inria.fr/algo/bsolve/constant/geom/geom.html

2. P. Kanerva, Sparse Distributed Memory, MIT Press.

MODULARITY AND MEREOLOGY

STEPHEN WOOD

Flat 3, 25 Lytton Grove, Putney, London SW15 2EZ

stephenwood60@hotmail.com

23/04/03

ABSTRACT

In the first section, 'Morphology and Classification', I give a brief account of the history of morphology and classification, examining the philosophical and theoretical basis of these disciplines. I criticise Darwin's contribution, questioning whether the explanations he gives in *The Origin of Species* are valid. I favour the cladistic theory of classification, which rejects the need for evolutionary assumptions.

In 'Genes and Modules', I reject neo-Darwinism's materialistic emphasis on genes. Nucleic acid genes rely on the infrastructure of the cell to carry out their role as replicators. The cell is a modular system of bounded compartments, on which the whole delicate balance of metabolism relies.

In 'Parts and Wholes', I contrast the mereology of Darwin with that of cladistics: historical groups comprising ancestor and descendants vs. relationships of organisms comprising common characters. These considerations of mereology lead on to a discussion of the appropriate theoretical basis of morphology and classification. In 'Specifications and Instances', I suggest such a basis is to be found in category theory.

1. MORPHOLOGY AND CLASSIFICATION

Pierre Belon, in his *La Nature des Oyseaux* (1555), places a skeleton of a bird alongside the skeleton of a human being. He shows that many of the bones in the two skeletons correspond and can be given the same names. This is a simply a pragmatic exercise, he says, to save time explaining the parts (I, ch. 12, p. 38). It is easier to discuss the anatomy of a bird, if we follow the same plan as in human anatomy (I, ch, 12, p. 39). The parts of the bird and the human correspond, but nothing is said of the relationship of birds and humans themselves. Belon is dealing solely with morphology—comparative anatomy—rather than classification.

Carolus Linnaeus, the great eighteenth century Swedish naturalist, is hailed as the father of modern classification. He describes six classes of animals, namely mammals, birds, batrachians (reptiles and amphibians), fishes, insects and worms. With Linnaeus, comparisons do not simply yield a pragmatic scheme; they reveal the inner design of nature itself. Each class is a rung on the ladder of perfection, the *scala naturae*, starting at the worms and rising to the highest of animals, Man. Birds carry some of the perfect characters of mammals, but not all. They correspond in many ways, but they occupy one rung lower down on the ladder than mammals.

The classification of animals without backbones has changed much since Linnaeus' time. The arrangement of backboned animals, excepting the split of the batrachians, has changed little until recently and remains the popular understanding. Fishes are cold-blooded, have scaly skins and lay eggs in the water. Amphibians live out of the water but must return to lay their eggs. Reptiles have a watertight skin and watertight eggs, and are thus able to live entirely on land. Birds and mammals are warm-blooded, birds being covered by a protective layer of feathers and mammals by fur. Birds have the power of flight. Most mammals bear live young that they continue to nourish by suckling. If we arrange nature as a scale of perfection, then it makes sense to classify living beings partly by the presence of characters, marking their perfection, and partly by the absence of characters, signifying imperfection. Thus, fishes are all backboned animals except amphibians, reptiles, birds and animals. They live permanently in water and do not have four limbs. They lack the characters of their increasingly perfect cousins.

Towards the end of the eighteenth century, Johann Wolfgang von Goethe applies a very different approach to living things. He looks not for perfection and imperfection, but for unity. He sees in individual plants, the One Plant that brings forth the many and yet remains always itself. The plants in their organs, leaf, stem and root, emerge as a multiplicity from the unity of the leaf. Looking one day on the shattered skull of a sheep, he sees a unity among the backboned animals: they are transformed vertebrae. Goethe's followers, Carl Gustav Carus (1828) and Richard Owen (1848) develop the type of all backboned animals (see Figure 1). The archetype is a series of vertebrae repeated along the long axis of the body. Simply, the unity of type is a common architectural plan, or *Bauplan*. Thus, we may recognise common elements in the type through their relative positions: 'The archetype skeleton represents the idea of a series of essentially similar segments succeeding each other in the axis of the body; such segments being composed of parts similar in

number and arrangement' (Owen, 1866: xiii). Geoffroy Saint-Hilaire (1818) understands the ear ossicles—the bones that lie in the middle ear in mammals, namely the malleus, incus and stapes—as the same as the opercular bones—the bones that cover the gills in fishes. Both lie behind the lower jaw and below the ear region of the braincase. Today, we would not accept Geoffroy's conclusion, but instead a similar one, made by Reichert in 1837. The ear ossicles correspond to bones at the back of the lower jaw, in fact lying beneath the opercular bones, namely the articular, the quadrate and the hyomandibular. Geoffroy espouses a principle of the constancy of relative position and connections. The generative process acts in such a way as to preserve the connections of elements invariant. A more subtle understanding of unity of type is therefore a unity of generative principle. Common plan manifests from common process, as an invariant. The generative principle is the source of multiplicity, not in a historical but in a rational, imaginative sense.

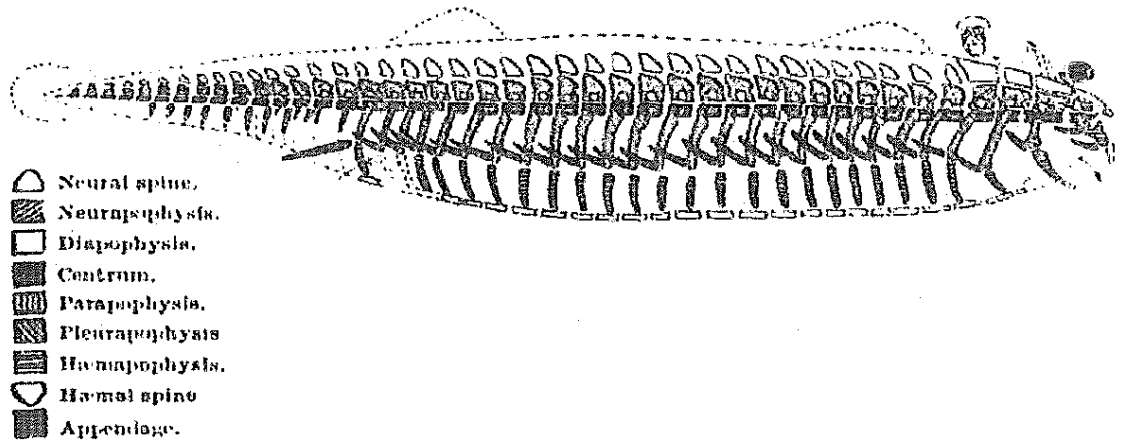


Figure 1: The vertebrate archetype from Owen (1848), constructed as a series of vertebrae.

Unity of type among living things of different modes of life is morphology's great claim. (The term 'morphology' owes to Goethe.) Darwin seeks to explain unity of type, in the thirteenth chapter of *The Origin of Species*. 'What can be more curious than that the hand of man, formed for grasping, that of a mole for digging, the leg of the horse, the paddle of the porpoise, and the wing of the bat, should be all constructed on the same pattern, and should include the same bones, in the same relative positions' (Darwin, 1859 [1968: 415]). The mole lives underground, the horse above it; the porpoise lives in the sea and the bat in the air. Yet, we may discern the same pattern of bones among these animals adapted to such different lifestyles. Darwin's explanation for the common plan lies not in a common generative principle but in a common historical origin. 'If we suppose that the ancient progenitor, the archetype as it may be called, of all mammals, had its limbs constructed on the

existing general pattern, for whatever purpose they served, we can at once perceive the plain signification of the homologous construction of the limbs throughout the whole class' (*loc. cit.*, p.416).

Morphological classification reveals a hierarchy of types within types, so for example, mammals are vertebrates and vertebrates are animals. Classification is 'the grand fact in natural history of the subordination of group under group, which, from its familiarity does not sufficiently strike us' (*loc cit.*, p.398). Darwin's explanation of classification is common descent: 'descent is the hidden bond of connexion which naturalists have sought' (*loc. cit.*, p.414). Humans, moles, horses, porpoises and bats are all descended from the common ancestral mammal, yet have all undergone modifications for adaptation to different modes of life. This is Darwin's theory of descent with modification caused by natural selection.

In a more modern classification of the backboned animals (Forey, pp. 30-31, pp. 128-143, in Burn, 1980), there are ten classes, six of which are devoted to fishes. The different types of fishes are as different from each other as they are from amphibians, reptiles, birds and mammals. The features that some fishes lack are striking. Hagfishes lack a fully developed backbone, instead the notochord, which is normally the embryonic precursor of the backbone, remains in the adult to form the main means of bodily support. (Indeed, hagfishes lie outside the backboned animals, and to include them we should talk of the craniates, animals with skulls.) Bone and cartilage are alien to the earliest fossil forms, the heterostracans. They have a different kind of skeletal tissue, aspidin, which lacks cells. Hagfishes, lampreys and the extinct heterostracans do not have jaws. Instead, hagfishes and lampreys have a muscular tongue with which they rasp at their prey, sucking blood. Neither do these three have paired fins, which arrive with another jawless group, the extinct cephalaspidomorphs. (For a fuller discussion, see Forey and Janvier, 1993, and the section on craniates at the Tree of Life website, http://tolweb.org/tree?group=Craniata&contgroup=Chordata).

The features that some fishes share with the tetrapods, or four-legged vertebrates, are just as surprising. Lungfishes cannot breathe in water. If they are held under water, they will drown. Young lungfishes have external gills, just like young amphibians. In the same way as amphibians, they lose their gills and gas exchange then takes place at the lungs. Lungfishes possess a double circulation, just like we do, blood from one half of the heart supplying the lungs, blood from the other half supplying the rest of the body. Look, next time you go to the aquarium at London Zoo. You may see a lungfish rise to the surface to breathe!

How have scientists arrived at the conclusions summarised above? Let's look at the modern process of classification. The data of classification are the characters. Examining a study group, the systematist will make a list of characters and their states, scoring the specimens in the study group as to which state they possess. Through some method, usually codified in a mathematical algorithm, such scorings are converted into the classification.

Willi Hennig's method is based on Darwin's theory of descent with modification: 'Evolution is a transformation of organisms in form and mode of life through which the descendants become different from their ancestors' (Zimmerman, quoted in Hennig, 1966: 88). Limbs are transformed fins; fins are primitive, yet limbs are derived. Fishes share the primitive character, fins, whereas tetrapods share the derived character. In Hennig's method, only the tetrapods are a genuine historical group; the fishes are not.

Consider the salmon, the lungfish and the cow. The cow may seem the most different, and the salmon and the lungfish most similar. However, as we have seen, the lungfish shares derived characters with the tetrapods that are not found in the salmon. The external gills and the two-chambered heart are the most striking. Other characters that the lungfish and the salmon have in common, such as paired fins not legs, and cold not warm blood, are primitive characters (see also Gee, 2000).

Hennig explicitly ties his method to a model of the evolutionary process. The derived characters of the lungfish and the cow are hypothesised to have appeared in their ancestral species. The lungfish shares a more recent common ancestor with the cow than with the salmon. The species ancestral to all three split to give rise to the salmon and to the ancestral species of the lungfish and the cow. 'Evolution in this sense (transformation) is also connected with speciation: if a species (reproductive community) is split into two mutually isolated communities of reproduction ... there is always a change (transformation) of at least one character of the ancestral species in at least one of the daughter species' (Hennig, 1966: 88). Hennig's method generates binary classifications: at every fork, two branches result. The two branches so formed are 'sister groups' (Hennig, 1966: 139).

For Hennig, morphology is not enough for classification. We need to add the tenet of descent with modification, of the evolutionary transformation of characters. Why are morphological correspondences—special homologies—not enough? Hennig answers that evolutionary

transformations involve the loss, not only the gain, of features. 'In general, we speak only of the homology of organs, but a "character" may also be the absence of an organ' (Hennig, 1966: 95). The absence of wings in silverfish and springtails is a shared primitive character, whereas the absence of wings in fleas and lice is shared derived. 'This cannot be expressed in an equally unequivocal way by saying that the absence of wings is a "special homology"' in the fleas and lice, but not in springtails and silverfish.

Fleas are 'a well-defined and homogeneous group', according to Davies (p. 146 in Burton, 1980). 'They are wingless, with a brown, bristly, strongly sclerotised body compressed from side to side. Their eyes are greatly reduced and the antennae are short but the legs are well developed and in many species the hind pair is used for jumping.' To be a flea is not simply to be wingless, but to have a compressed body, short antennae and well-developed legs. We classify them together because of these positive characters. Fleas undergo complete metamorphosis (holometaboly), where they pass through a pupal stage between larva and adult. In the pupa, the larval tissues die and break down, and the adult tissues grow from special zones of persistently embryonic tissues, known as the imaginal discs. Since most insects that completely metamorphose have wings, then it is likely that fleas have lost their wings. The position of fleas within the winged insects is based on characters they possess, not on characters they lack. The fact that fleas lack wings just makes the business of classification more frustrating.

Norman Platnick applies a similar argument to the absence of limbs in fishes: 'if we form a group Pisces, we have based it not on a character, but on the absence of a character. The group Pisces includes those organisms with fins that also happen to lack modified fins (limbs). Such use of the absence of a character is one of the hallmarks of an artificial group' (Platnick, 1980: 544). Paired appendages unite the cephalospidomorph fishes with the gnathostomes (backboned animals with jaws). Tetrapods have these paired appendages manifesting as limbs and are one type of gnathostome. Sharks, lungfishes and codfishes are others. There is level in the hierarchy of types where the presence of paired appendages is relevant: allying the cephalaspidomorphs with gnathostomes and not, for example, the lampreys or the hagfish. However, with the Cephalaspidomorpha + Gnathostomata, we cannot pull together as Pisces all those animals which have paired appendages but not limbs. In this context, the presence of paired appendages is no longer relevant. Colin Patterson (1982) points that every shared primitive

character is just a shared derived character considered at the wrong hierarchical level.

Morphology is the study of the affinities between parts of organisms; classification is the study of the affinities between organisms as a whole. According to Hennig, we bridge the gap by considering transformations between corresponding parts. According to Patterson, we bridge the gap simply by considering the hierarchical level of correspondences. Every homology is a shared derived character somewhere in the hierarchy of life. We either assume a *process* of transformation, or a hierarchically branching *pattern*. Cladistics is the study of branching diagrams (Nelson 1979) and the approach of Nelson, Platnick and Patterson is known as pattern cladistics.

2. GENES AND MODULES

There are difficulties in morphological classification, in the delimitation and correlation of characters, which have led some systematists to genetics instead. Where does one character start and another stop? Do we consider a suite of features, which make up a coherent functional whole, one character or many? What are the dependencies between the states of one character and the states of another? These correlations could be functional or pleiotropic. For example, all the different adaptations to endothermy ("warm-bloodedness") in birds and mammals are functionally correlated, whereas a mutation that causes both blindness and a deformity of the skeleton in mice is known as a pleiotropic correlation.

Much of the literature on morphological classification deals with disagreements over the conceptualisation of characters. Often such disagreements arise because living things vary continuously rather than discretely. Many of the algorithms employed for morphological classification require discrete characters. So, systematists need to find a way to divide the range of continuous variation into a number of relatively distinct states. Also, the processes of morphological change are only just becoming understood. How can we classify when we know so little about the possibilities of character state transformation? For example, we may reason that it is unlikely that lice and fleas are unlikely to be related by the loss of wings, because in evolution it is easier to destroy than to create (Mayr, 1942: 278-279). However, it would be better if we had a testable model of morphological change to back up such reasoning.

What about gene sequences? Do we have more success here? The base pair is the character and it is possible to construct tractable, statistical models of genetic change. The model can take account of dependencies between characters. For example, base pairs in the first codon position are more likely to be constrained by function than base pairs in the third position. This is because the base pair in the third position often makes no difference to the aminoacid into the ribosomes translate the gene sequence.
Base pairs are the units of replication, mutation and recombination, the material cause of genetic evolution. They are the hereditary atoms that are passed on from generation to generation, ensuring material continuity. In this mechanistic approach, statistical models of genetic change represent the efficient cause of evolution. Models may simply describe a certain probability for base pairs to be miscopied. (Imagine a machine running down.) Alternatively, models can be increased in complexity by considering the external forces of natural selection. (Imagine billiard balls pushed here and there.) So to avoid the problems with delimitation and correlation of characters we may turn to a material approach.

In Darwin's account, form is reduced to matter. Form is preserved from generation by generation, through the material connection of descent. Darwin does not say why form should be preserved from a common ancestor, and this seems a little at odds with 'descent with modification'. To explain the unity of type, we have to assume 'descent *without* modification'! If we bring natural selection into the equation, then we can say that form is preserved as long as it is not disadvantageous and removed by selection. So instead of 'descent with modification caused by natural selection', we have 'descent without modification caused by the absence of natural rejection'. Rather than explaining morphology and classification, Darwin explains them away!

We may have doubts about the present inability of a genetic approach to account for all the subtleties of morphology. Is it only a matter of time before genetics fully explains the riddles of organic form, or is there some fundamental impediment to its progress? Today all life is based on nucleic acid, but life could not have started that way. Graham Cairns Smith (1982) raises 19 objections to the synthesis of nucleic acids in the early oceans. There are difficulties with the purity, the concentration and the instability of intermediates. The replication of nucleic acids requires processes and structures that are too complex to take place outside a living cell: intricate metabolic pathways, delicately synthesised enzymes, organised structures such as ribosomes. Much more likely, life started with materials that were common in the early Earth. The origin of life

was 'low-tech' rather than 'hi-tech', unlikely to have been based on sophisticated, manufactured components.

Cairns Smith's controversial suggestion is that life started as clay crystals—replicators that were not DNA. Clay crystal layers grow one on top of another. Dislocations propagate through the crystal lattice like mutations in a lineage of bacteria. Some time after the origin of life, there was a genetic takeover, where DNA replaced clay in its role as replicator. All today's life is likely to have been descended from a common ancestor with nucleic acid as the replicator. This does not mean that all life that has ever lived replicated through nucleic acid.

Apart from being the first replicators, clay crystals would also have had to provide the first cellular morphologies. Clays are inert, not readily reacting with organic molecules in aqueous conditions. They would not have interfered with complex, drawn-out synthesis of organic molecules. Clays can form the structured compartments necessary to concentrate reactants and to purify products in a metabolic pathway. Halloysites form membranous tubes and vesicles. Zeolites and sepiolites pack together as bundles of grooved rods, with well-defined pores.

Cells and their compartments must have been present to allow DNA replicators to evolve. 'These complex structures, we are beginning to suspect, hold much of the secret of how cellular processes are controlled in both space and time: the secret may consist, at least in part, of isolating and maintaining the different cellular constituents—mainly enzymes, together with their substrates, products and modifiers (activators, inhibitors)—in different compartments; sometimes allowing, sometimes denying mutual accessibility' (Mahler and Cordes, 1971). Clays would have formed the 'glassware' for early organic reactions.

The neo-Darwinian definition of life is to replicate with error: 'We shall regard as alive any populations of entities which has the properties of multiplication, heredity and variation' (Maynard Smith, 1975). However, this cannot be the whole story. Living cells are bounded and themselves contain bounded components. This is vital to metabolism. Living cells are stable to resist error catastrophe, forestalling the build-up of deleterious errors. At the same time, favourable errors are accumulated and preserved over generations. Living cells are autonomous, bounded, stable entities, creating and maintaining the boundary between themselves and their surrounding world. 'Animals and plants are made out of homologous organelles like the mitochondria, homologous organs like the gills and lungs, homologous limbs such as

arms and wings. They are the stable holons in the evolutionary flux' (Koestler, 1967: 139).

Life is not simply a replicating system; it is an organised, modular system. Living organisation is the end of 'an ordered and structured inner movement that is essential to what things are' (Bohm, 1980: 12). Far from being explained away by theories of genetic descent, morphology, with the cell as the unit of morphology, lies at the centre of a theory of life, not DNA.

3. PARTS AND WHOLES

In the Darwinian picture, an ancestor and its descendants form a whole in the manner of an individual, with a unique origin, history and termination. The taxa discovered through classification are wholes, with organisms as their parts. If ancestry is not considered, taxa lose their claims to individuality and classifications are nothing more than diagrams of set membership (Friday, 1994). The taxa that emerge in a classification are, therefore, accidents of history. Species descended from a common ancestor form a historical group. There is no necessary—or rational—connection between them. Without common descent, the hierarchy of types is nothing more than an arbitrary nesting of sets. Ho (1990) and Goodwin (1994) complain that Darwin thus removes the scientific status of biology by replacing rational theories of biological organisation with purely historical narratives. Darwin changes the language of biology, away from talking about types—common generative principles—to historical groups of ancestral and descendant species.

Gary Nelson considers an alternative picture to Darwin. Taxa are not groups of organisms, but relationships. 'Conceived as relationships, taxa and homologies do not literally descend from one another. Taxa come into being with organisms that literally descend' (Nelson, 1989: 281)'. The unity of the tetrapod type, for example, consists in the homologies, or morphological correspondences, that tetrapods share. To justify the hypothesis of the tetrapod type, there is no need to propose that there existed a tetrapod ancestor, which possessed these features. No ancestral animals have been found and many evolutionary arguments are advanced to account for this. For example, if new species emerge in small, isolated populations, then it is unlikely that any members of the new species will die in circumstances suitable for their remains to be fossilised. The validity of the Tetrapoda does not depend on the ancestral tetrapod being found. To recognise a morphological correspondence is to make a

hypothesis, not about ancestry, but about the level in the hierarchy of life at which it applies. Every homology is part of some taxon, contributes to the unity of some type.

A horseshoe bat and I are organisms, and one of the parts of my body, my arm, is a homologue of the bat's wing. Following Darwin, we would say that the bat and I *belong* to the taxon Tetrapoda, as descendants of the ancestral tetrapod. However, Nelson says that the bat and I are *related* through the taxon Tetrapoda, and our arms/wings are related by the homology 'limbs'. This homology is part of the Tetrapoda. 'Taxa are homologies and have homologies for their parts. Organisms are homologues and have homologues for their parts' (Nelson, 1989: 279). Nelson describes homologies and taxa as 'phylogenetic parts of life', as opposed to homologues and organisms, which are 'ontogenetic parts of life'. Phylogenetic parts of life expose the hierarchical pattern of life, the pattern of common descent, if you will. Ontogenetic parts of life express the history of individual living things.

There are a number of subtle ideas here, which we will need to explore further:

1. The distinction between phylogenetic and ontogenetic parts of life, and the relationship between the two.
2. The distinction between homologies and taxa and the part-whole relationship between the two.
3. The relationship between taxa at different hierarchical levels.

All these ideas derive from Patterson (1982). To recognise a morphological correspondence is to make a hypothesis about the level in the hierarchy of life at which it applies, and thus to make a hypothesis of a taxon. My Ph.D. supervisor, Ken Joysey, commissioned this paper from Patterson. He hoped that, forced to consider the question of homology, Patterson would give up the pattern approach he had adopted from Nelson and Platnick. Instead, Joysey relates, 'Patterson gave up evolution'! In the same symposium volume, Alan Charig comments that, without evolution, pattern cladistics 'is without any proper theoretical basis' (Charig, 1982: 372). Statistics underpins genetic approaches to reconstructing evolutionary history. The likelihood of and support for a particular tree are well defined within the theory. The suggestion I make in this paper is that the appropriate mathematical language for cladistics is *category theory*, and in particular, colimits. I examine the application of category theory to hierarchies in computer science and explore its relevance to systematics.

4. SPECIFICATIONS AND INSTANCES

'How do you build big, complicated things? Commonly, you put them together—or *configure* them—from smaller, simpler things, and this is as familiar in software development as it is in everyday life' (Vickers and Hill, 2001: 32). This is Herbert Simon's message (Simon, 1962), that the architecture of complexity—natural and artificial—is modular. Arthur Koestler (1967) draws extensively on Simon in his hierarchical theory of holons, which he applies to behaviour, language and morphology.

Software modules can be reused in contexts different to the one in which they originated. The internal implementation of the module can change as long as the services that it offers remain the same. Enhancements and bug fixes are limited in scope to one or a few modules. These are the principles of object- and component-oriented programming. They are related to Simon's insight, that modular systems are more stable. Subsystems may change independently of one another, without affecting the integrity of the system.

The object-oriented programmer defines the services offered by a module as a class. The properties will generally be held internally, invisible to the outside world, and methods will be available to get and set these properties. A running program will create objects as instances of the class. Each object may have different data values, but conforms to the type of module specified in the class.

Steven Vickers and Gillian Hill address the problem of putting together systems from specified components and specified sharing. They contrast specifications with their instances. This is a very natural language to talk about classification. Taxa are specifications and organisms are instances. The horseshoe bat and I are instances of the tetrapod specification, which is composed of the homology 'four limbs'. This defines the tetrapod type. The horseshoe bat and I possess instances of four limbs as parts of our body. Instances of homologies are homologues. Nelson's phylogenetic parts of life are specifications; his ontogenetic parts of life are instances.

In a similar vein, Stanley Salthe reflects on the hierarchy of classification and concludes that it is a hierarchy of specifications. Hierarchies, say in ecology, are usually considered in scalar terms, where 'components at different levels differ in size roughly by orders of magnitude ... Levels in the specification hierarchy mark the qualitative differences of different realms of being, as in 'physical realm' versus 'biological realm'' (Salthe,

2001, http://www.nbi.dk/~natphil/salthe/Hierarchy_th.html). One of the tasks of the systematist, when presented with a specimen, is to give as precise a specification of its identity as possible. With fossils, which may not be complete, the identification may be carried only to family level, leaving the genus unspecified, or to genus, leaving the species unspecified. Levels in the specification hierarchy emerge in a generative process, starting with the most general and leading to the most specific. At each stage, the range of accessible states narrows irreversibly (Salthe, 1993: 64-65).

So how does category theory fit in? A category is 'a mathematical universe', populated with certain objects, maps between these objects, and rules for how the maps can be added together (Lawvere and Schanuel, 1998). Maps between objects are often called morphisms.

Consider specifications, X and Y, to be objects in a category C: '... a morphism $u: X \dashrightarrow Y$ describes how the specification X can be considered a part of Y. In consequence, an implementation—or *instance*—of Y contains somewhere inside itself an instance of X, and we shall view this *reduction* of instances of Y to those of X as the most direct manifestation of morphisms. (Note the contravariance. The natural direction of morphisms for instances is the reverse of the natural direction for the syntax of specifications.)' (Vickers and Hill, 2001: 32)

So, in Nelson's category of life, specifications X and Y are phylogenetic parts of life, namely homologies and taxa respectively. The implementations, or instances, of X and Y are homologues and organisms. The map u assigns a homology X to the taxon Y of which it is a part. In consequence, an instance of taxon Y contains within itself an instance of the homology X. For example, *u: four limbs --> Tetrapoda* means that tetrapod organisms possess homologues of four limbs, whether arms, legs, wings or flippers. The language of specifications and instances, framed in category theory, captures the subtlety of Nelson's mereology.

'Now suppose Y_1 and Y_2 both contain X, by morphisms $u_i: X \dashrightarrow Y_i$. Then the pushout Z, a simple colimit, configures a new specification out of Y_1 and Y_2 with sharing a common part X' (Vickers and Hill, 2001: 33). Let's consider a concrete example (Vickers and Hill, 2001: 43-44). *Lift* is the class of instances of lifts in an office building. *CButton* is the class whose instances are series of call buttons. There is an instance reduction from *Lift* instances to *CButton* instances. Given any lift, we can point out its call buttons and so display a *CButton* instance. The category C of primitive specifications includes the objects *Lift* and *CButton* and a map

i: CButton --> Lift. If L_1 is a lift instance, then i_1L_1 will denote its call buttons. We would like to specify a *LiftSystem* with two lifts serving the same location. Naturally, they should share the same call buttons. We have two lifts instances L_1 and L_2 and a set of call buttons B, which is common, i.e. $i_1L_1 = i_2L_2=B$. Identifying the pattern of call buttons common to the two lifts allows us to recognise a specification hierarchy. The lifts are at level n in the hierarchy, the lift system at level $n+1$.

This is exactly Patterson's insight that the hypothesis of a morphological correspondence between instances of two taxa is equivalent to the hypothesis of a taxon at a higher level. Indeed, the two taxa are sister taxa within the newly discovered taxon.

Consider an example where there are three taxa, hagfishes, lampreys and gnathostomes, or jawed vertebrates (see Figure 2, below). Hagfishes have a skull, but no backbone or jaws. Lampreys have a skull, a backbone, but no jaws. Gnathostomes have a skull, a backbone and a pair of jaws. The homology 'backbone' shows that lampreys and gnathostomes are sister taxa. This relation identifies the Vertebrata, which emerges as the colimit. In turn, the homology 'skull' shows that vertebrates and hagfishes are sister taxa, identifying the Craniata as a further colimit. Hierarchy bridges the gap between the morphology and classification and emerges naturally from the theory of categories.

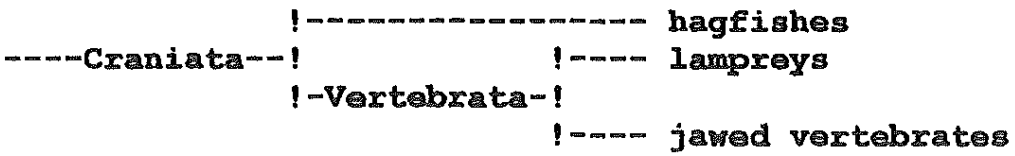


Figure 2: The classification of craniates.

So far, the algorithms employed in cladistic analysis require discrete characters. We have already mentioned that this raises a difficulty with continuous characters, which have to be split up into a number of discrete states. Ideally, data should be entered into the analysis in as raw a form as possible. Categories are abstract, wherein lies their power. If homology can be formalised in the way that the above discussion suggests, then we have a concept that applies to continuous characters as equally as to discrete characters. I am hopeful that new methods of cladistic analysis will emerge from the application of category theory.

ACKNOWLEDGMENTS

Thanks to Keith Bowden for discussions of hierarchies, categories and presheaves, which led to my discovery of the work by Vickers and Hill.

REFERENCES

Belon, Pierre (1555) *L'histoire de la nature des oyseaux.* Facsimile reprint with introduction and notes by Philippe Glardon. Droz, Geneva, 1997

Bohm, D. (1980) *Wholeness and the Implicate Order.* Routledge and Kegan Paul, London.

Burn, David M. (1980, ed.) *The Colour Encyclopaedia of Animal Life.* Peerage, London.

Burton, Maurice (1980, ed.) *The New Larousse Encyclopaedia of Animal Life.* Hamlyn, London.

Cairns-Smith, A. G. (1982) *Genetic Takeover.* Cambridge University Press, Cambridge.

Carus, Carl Gustav (1828). *Von der Ur-Theilen des Knochen- und Schalengerüstes.* Leipzig.

Charig, A. J. (1982) Systematics in biology: a fundamental comparison of some major schools of thought. In Joysey, K. A. and Friday, A. E. (eds.) *Problems of Phylogenetic Reconstruction,* pp. 363-340. Academic Press, London.

Darwin, C., 1859[1968] *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* Murray, London. Reprinted by Penguin, London, edited with an introduction by J. W. Burrows.

Forey, P. L., and Janvier, P. (1993). Agnathans and the origin of jawed vertebrates. *Nature* 361: 129-134.

Friday, A. E. (1994) Adaptation and phylogenetic inference. In: Eggleton, P. amd Vane-Wright, R. I. (eds.) *Phylogenetics and Ecology,* pp. 207-217. Academic Press, London.

Gee, Henry (2000). *Deep Time.* Fourth Estate, London.

Geoffroy Saint-Hilaire, E. (1818) *Philosophie Anatomique.* Paris.

Goodwin, B. [C.] (1994) *How the Leopard Changed its Spots: The Evolution of Complexity.* Weidenfeld and Nicholson, London.

Hennig, W. (1966) *Phylogenetic Systematics.* University of Illinois Press, Urbana.

Ho, M.-W. (1990) An exercise in rational taxonomy. *Journal of Theoretical Biology* 147: 43-57.

Koestler, A. (1967) *The Ghost in the Machine.* Hutchinson, London.

Lawvere, F. W. and Schanuel, S. H. (1998) *Conceptual Mathematics: A first introduction to categories*. Cambridge University Press, Cambridge.

Mahler, H. R. and Cordes, E. H. (1971) *Biological Chemistry*. 2nd edn. Harper and Row, New York.

Maynard Smith, J. (1975) *The Theory of Evolution*. 3rd edition. Penguin, Harmondsworth.

Mayr, E. (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

Nelson, G. J. (1979) Cladistic analysis and synthesis: principles and definitions with a historical note on Adanson's *Familles des Plantes* (1763-1764). *Systematic Zoology* 28(1): 1-21.

Nelson, G. J. (1989) Cladistics and evolutionary models. *Cladistics* 5: 275-289.

Owen, R., 1848 *On the Archetype and Homologies of the Vertebrate Skeleton*. London.

Owen, R., 1866 *On the Anatomy of Vertebrates*, Vols. I and II. London.

Patterson, C., 1982 Morphological characters and homology. In Joysey, K. A. and Friday, A. E. (eds.) *Problems of Phylogenetic Reconstruction*, pp. 21-74. Academic Press, London.

Platnick, N. I. (1980) Philosophy and the transformation of cladistics. *Systematic Zoology* 28: 537-546.

Salthe, S. N. (1993) *Development and Evolution: Complexity and Change in Biology*. MIT Press, Cambridge, Massachussetts, and London.

Simon, H. A. (1962) The architecture of complexity. *Proceedings of the American Philosophical Society*. 106: 467-482.

Vickers, S. and Hill, G. (2001) Presheaves as configured specifications. *Formal Aspects of Computing* 13: 32-49.

The Topos of Emergence

or: putting the tower of turtles on a more *profound* basis

or: more about objects which are not even intended to be entities

by DAN KURTH

Figures by VERA HOHLSTEIN

Institut für Geschichte der Naturwissenschaften
Johann Wolfgang Goethe Universität
60054 Frankfurt am Main
Germany
e-mail: dankurth@web.de

Abstract. The aim of this paper is to provide a sketch of a bit more formal underpinning concerning the conceptual intricacy of the notion of '(primordial) emergence' by applying category- and topos-theoretical means.[1] This relates most directly to what had been called 'the mêontology of primordial emergence' in the corresponding paper "The Tower of Turtles" (see the "The Philosophical Aspects of ANPA 24"), namely the 'objects which are not intended entities'.

A particularly interesting instance of such 'objects which are no entities' is the proposed mathematical structure from which in my understanding the most elementary (or fundamental) physical structure(s) have been emergent.

1 Introduction

In this paper[2] I will try to give an outline of a mathematical model of 'primordial emergence'. By 'primordial emergence' is meant an emergence of primordial physicality up from an unphysical *purely intelligible* mathematical state. Intimately intertwined with that I will also present a sketch of a mathematical structure apt to solve a seeming paradox of the reductionist program in science which I've discussed in the directly related paper 'The Tower of Turtles'[3] (see the volume containing 'The Philosophical Aspects of ANPA 24') where I stressed the point that the reductionist program cannot come to its very end, i.e. it cannot terminate as long as it is pursued as a project with the aim of discovering a primordial and fundamental level of physical reality which doesn't stem from any other (non-physical) preceding structure.

---

[1] The idea of applying features of topos-theory as means for analyzing the concept of emergence and subsequently formulating a theory of emergence had - to my knowledge - for the first time been proposed in P.Eisenhardt, D.Kurth, Emergenz und Dynamik, Cuxhaven 1993

[2] This paper is based on the talk I gave at the ANPA 24 meeting in Cambridge (UK), August 2002

[3] D.Kurth, The Tower of Turtles, in: The Philosophical Aspects of ANPA 24, Proceedings of ANPA 24, to appear

I.e. I did state in this paper that it is a hidden (yet perhaps mostly unnoticed) implication of the reductionist project that physical existence itself has to be seen as being emergent from an underlying pre-physical level, which I would call a 'mêontological' level, by this indicating that this level has to be seen as being made of objects which are no entities and which cannot even be intended to be entities, i.e. a level of *not in any sense physically existing objects*. Thus these objects should - since I'm not in favour of spiritual entities (at least not in this context) - be *mathematical objects*.
As an inspiration for such kind of objects I had a look into somewhat related considerations of Leibniz where I found a sketch of something I've called 'dynamical Leibniz-point objects'.[4]
The means I now will use to give the Leibniz-point objects a modern spin are mainly the ones you might expect. Namely

a) n-categories or higher dimensional categorical algebra[5] (yet in our case it will turn out to be rather lower dimensional, i.e. negative ascending categorical algebra) and intricately intertwined with that

b) Topos theory

But before we will try to make a suggestion of how to bring such reconstructed Leibniz-point objects (despite the fact that they shall not be intended to be entities) into appearance, we must at first say a word about our interpretation of the *conatus* which by Leibniz was meant to be copresent with his special kind of points (i.e. points which have parts (or a structure) yet no extension). Leibniz obviously took the conatus as an infimum motus, i.e. as an infinitesimal motion and as such of course as an infinitesimal physical entity.
Such an infinitesimal physical entity could obviously not serve for our purpose of overcoming the difficulties of reductionism and therefore we will give this dynamical aspect attached to that unconventional Leibniz point object a different meaning. Like the Leibniz point object itself we also will understand the conatus as an entirely mathematical object, i.e. as something *dynamical in a purely mathematical sense*. There is already a tradition in doing so and taking in particular morphisms as formal analogies of dynamical action.[6] Thus we will suggest to see the conatus as an automorphism of a Leibniz point object onto itself. And motivated by this new aspect we now will also rename our objects in question and instead of 'Leibniz point objects' we will just call them 'automorphic objects' ($\mathbf{ob_{AM}}$).
Now two other preliminary terminological remarks (yet this time not about a terminology of our own making). The mathematical concept of an n-category had been developed in two different mathematical contexts and there are also two different types of n-categories, namely strict n-categories and weak n-categories. Since we will later make use of both these types let me give now a short characterisation.
Strict n-categories are the much older and in a sense less fruitful bunch. Yet they played an important role in metamathematical considerations e.g. in the mid sixties when they had been used by F.W.Lawvere[7] in the business of foundational efforts. Lawvere's n-category (the category of categories **CAT**) then served a metamathematical purpose. That kind of n-

[4] For more details cf. D.Kurth, The Tower of Turtles, loc. cit.
[5] The idea of applying n-categorical concepts and means for analyzing emergence has already been introduced (and extensively discussed) in: P.Eisenhardt, D.Kurth, Complexity Categorified; in: Implications, Scientific Aspects of ANPA 22, (ed. Keith Bowden), London 2001
[6] Cf. as an example F.W.Lawvere, Categorical Dynamics, in: Topos theoretic methods in geometry (ed. A.Kock et. al.), pp. 1-28. Aarhus University Matematik Institut various publication series no. 30, Aarhus 1979
[7] Cf. F.W.Lawvere, The category of categories as a foundation for mathematics, in: S.Eilenberg et. al. (eds.), Proceedings of the Conference on Categorical Algebra in La Jolla, 1965

categories later had been named strict *n*-categories (i.e. a kind of *n*-category in which the specific content or structure of the *n*-1 category is not so much of a particular relevance) in difference to the so called weak n-categories which are no metamathematical products of foundational endeavours yet categorical representations of increasingly enriched layers of hierarchical structures. I.e. in a *strict* *n*-category the object of the strict *n*-category is the respective *n*-1 category itself whereas in a *weak* *n*-category the object(s) of the weak *n*-category is (are) the *morphism(s)* of the respective *n*-1 category. By J.Baez, an important contributor to the development of higher dimensional categorical algebra (read as '*n*-category theory'), the definition and the significance of weak *n*-categories has been put into the following words

> "An *n*-category is an algebraic structure consisting of a collection of 'objects', a collection of 'morphisms' between objects, a collection of '2- morphisms' between morphisms, and so on up to *n*-morphisms, with various reasonable ways of composing these *j*-morphisms. A 0-category is just a set, while a 1-category is just a category. Recently *n*-categories for arbitrarily large *n* have begun to play an increasingly important role in many subjects including homotopy theory and topological quantum field theory. The reason is that they let us *avoid mistaking isomorphism for equality*."[8]

The mathematical concept or object of a *topos* had been originally designed by William F. Lawvere and Michael Tierney to be a dynamical version of a set.[9] By dynamical here is meant that there are (non-trivial) morphisms between any of the objects of the topos , where these objects are roughly speaking the categorical version of the elements of set theory. Further characteristics of a topos are amongst others that any topos has to have an initial object as well as a terminal object. The mentioned dynamical aspect then turns out to be equivalent to the various sequences of morphisms which can be found in the universe between the initial and the terminal object.
In the following we will make use of all three concepts we've just touched. And by making use of them we can maybe also give an additional philosophical spin to the difference of 'impact' related to strict respectively weak *n*-categories.

2 The topos of pre-physical emergence (PrePhys): automorphic objects with a (negative ascending) n-categorical enrichment

The following figure FIG.1 shows the mathematical structure I propose as being fit to carry the burden of the tower of turtles or being apt to stand as a profound basis for it. FIG.1 shows the topos **PrePhys** which is meant to represent an assumed pre-physical as well as pre-natural mêontological 'process' and that process itself is characterised by a strict *n*-categorical unfolding of automorphic objects $\mathbf{ob}_{\mathbf{AM}}$.
So we mix toposic and *n*-categorical features in **PrePhys**. And by speaking of a 'mêontological process' we do even something more strange than that. The notion of a 'mêontological process' implies a process of (or in) the non-being.
Thus it must not come as a surprise that the automorphic objects $\mathbf{ob}_{\mathbf{AM}}$, which are the particular sort of subobjects of **PrePhys**, are exactly the candidates for the objects which are not and cannot even be intended to be entities but which are intended to be the matrix from

[8] Baez, J., An Introduction to *n*-Categories, in. 7th conference on Category Theory and Computer Science, ed. E.Moggi and G.Rosolini, Springer Lecture Notes in Computer Science vol. 1290 (1997) p 1
[9] Cf. F.W.Lawvere, An elementary theory of the category of sets, Proceedings of the National Academy of Sciences, U.S.A., 52, 1506-11; F.W.Lawvere, Continuously variable sets: algebraic geometry = geometric logic, in: Proceedings of the ASL Logic Colloquium, Bristol 1973 (ed. H.Rose, J.C.Shepherdson), pp. 135-56

which such objects can arise (or emerge) which then could justifiably be intended to be entities.[10] To shed some light on such peculiar objects has been the motivation and purpose of this paper as had been explained in the introduction.

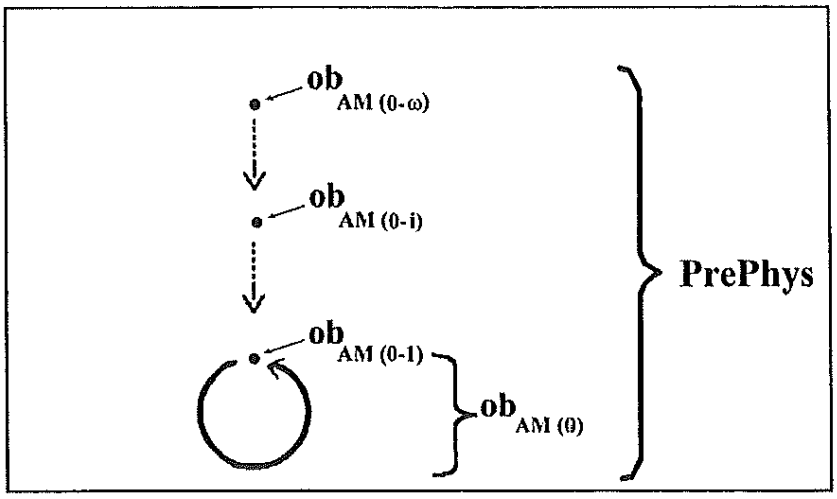


FIG.1: The topos of pre-physical emergence **PrePhys**: automorphic objects with a (negative ascending) strict *n*-categorical enrichment

For **PrePhys** the following definitions hold:

PrePhys: $\{\mathbf{ob}_{AM(0\text{-}\omega)}, \mathbf{ob}_{AM(0\text{-}i)}, \mathbf{ob}_{AM(0)}\}$

with $\mathbf{ob}_{AM(0\text{-}\omega)}$ (or a respective $\mathbf{ob}_{AM(0\text{-}i)}$) as initial object and $\mathbf{ob}_{AM(0)}$ as terminal object
for $1 \leq i \leq \omega$
and with $\omega >> n$, $\omega >> n+1$, $\omega >> n+2$, ..., $\omega >> n+j$, $\omega >> n+k$, ...;
with $k >> j$ (for $n, j, k \in \mathbb{N}$)
ω then is an infinitely large Natural Number, and thus it holds $\omega \neq \infty$.

Any of the automorphic objects $\mathbf{ob}_{AM}$ which are the subobjects of **PrePhys** then is itself also a category, consisting of an object $\mathbf{ob}_{AM}$ and an (auto)morphism $\mathbf{am}_{obAM}$. Yet it must be noted that the very structure of **PrePhys**, i.e. the relations between its subobjects $\{\mathbf{ob}_{AM(0\text{-}\omega)}, \mathbf{ob}_{AM(0\text{-}i)}, \mathbf{ob}_{AM(0)}\}$ is a strict *n*-categorical unfolding. Therefore it holds for arbitrary subobjects $\mathbf{ob}_{AM}$ of **PrePhys**:

a)Any subobject $\mathbf{ob}_{AM}$ of **PrePhys** has an *n*-categorical double $\mathbf{ob}_{AM(-i)}$.
b)An *n*-category $\mathbf{ob}_{AM(-i)}$ with the rank (or dimension) -i consists of an object $\mathbf{ob}_{AM(-i-1)}$ and the automorphism $\mathbf{am}_{obAM(-i-1)}$ onto that object $\mathbf{ob}_{AM(-i-1)}$ itself.
(For $0 \leq i \leq \omega$; $i \in \mathbb{N}$ and ω being an infinitely large Natural Number)

As a further illustration of that *n*-categorical aspect of the internal making of **PrePhys** might serve FIG.2, which shows $\mathbf{ob}_{AM(0)}$ and its internal structure consisting of $\mathbf{ob}_{AM(0\text{-}1)}$ and the automorphism $\mathbf{am}_{ob(0\text{-}1)}$ of $\mathbf{ob}_{AM(0\text{-}1)}$ onto itself in some magnification.

[10] For all the anti-epistemological, ontological and mêontological intricacies which had been implied in these remarks cf. D.Kurth, Actual Existence and Factual Objectivation, in: Movements, Philosophical Aspects of ANPA 23 (Proceedings of ANPA 23), Arleta D. Ford (ed.), London 2002 and D.Kurth, The Tower of Turtles, in: The Philosophical Aspects of ANPA 24, Proceedings of ANPA 24, to appear

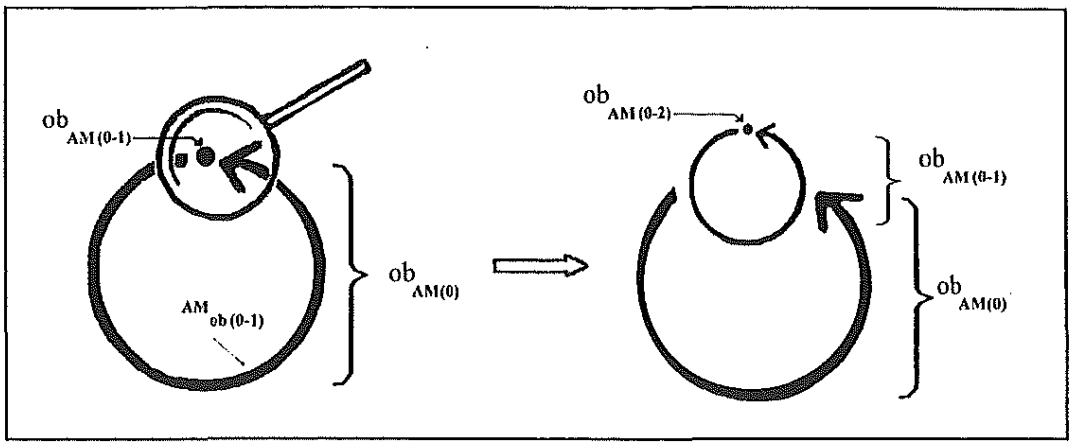


FIG.2: $\mathbf{ob}_{AM(0)}$,$\mathbf{ob}_{AM(0-1)}$ and $\mathbf{am}_{ob(0-1)}$ as seen through a magnifying glass

The following figure FIG.3 then might give the reader a broader perspective on that mentioned mêontological process, which can be characterised as a potentially infinite regress into the entrails of the pre-physical mêon, the non-being, itself. And it might be a question of terminological taste whether to call that process an 'unfolding' or an 'enfolding'.

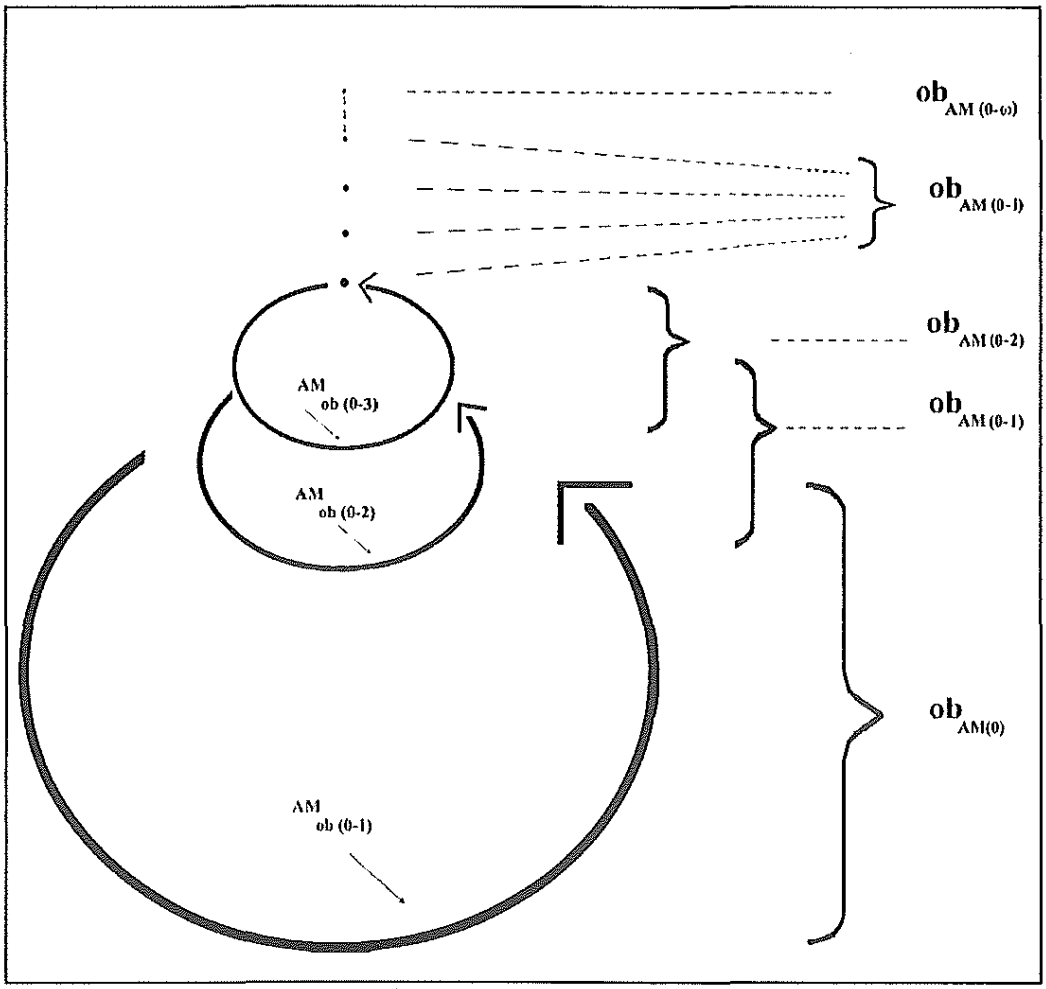


FIG.3: The potentially infinite involutive regress in **PrePhys**: a negative ascending inverted hierarchy

PrePhys has the automorphic object $\mathbf{ob}_{AM(0-\omega)}$ or a respective $\mathbf{ob}_{AM(0-i)}$ as its initial object and the automorphic object $\mathbf{ob}_{AM(0)}$ as its terminal object, the functors between the various $\mathbf{ob}_{AM(0-i)}$ (possibly *up from* $\mathbf{ob}_{AM(0-\omega)}$) terminating at $\mathbf{ob}_{AM(0)}$ all being isomorphisms.
The automorphic object $\mathbf{ob}_{AM(0)}$ and the various automorphic objects $\mathbf{ob}_{AM(0-i)}$ (possibly including $\mathbf{ob}_{AM(0-\omega)}$) then are the subobjects of **PrePhys**.

In addition - and even somewhat in contrast - to these topos-theoretical characterisations there are also the respective *n*-categorical features namely that the objects $\mathbf{ob}_{\mathbf{AM(0\text{-}i)}}$ signify a (negatively ascending) hierarchy of respectively lower (or rather 'negatively higher') dimensional strict *n*-categories (possibly *up to* $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$) related to the one from which this 'descending' (i.e. negatively ascending) cascade starts namely $\mathbf{ob}_{\mathbf{AM(0)}}$. Thus it must be noted that even though $\mathbf{ob}_{\mathbf{AM(0)}}$ is - so to speak - the starting point of a *construction* of negatively increasing categorical dimension it very well also is - perhaps at the first look somewhat puzzling - the object with the *highest* dimension in the particular hierarchy produced by an assumed underlying *process* - simply due to the fact that $0 > 0\text{-}i$ (for $i \in \mathbb{N}$).
The mentioned contrast of the topos-theoretical characterisation of **PrePhys** and the characterisation of the *n*-categorical construction of the $\mathbf{ob}_{\mathbf{AM(0\text{-}i)}}$ (possibly *up to* $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$) (which also happen to be the subobjects of **PrePhys**) has - in my view - to do with two problems. Namely a) with the difference between a process and a construction, and then subsequently b) with the problem of how to envision a pre-physical or pre-natural process.
Concerning a) it seems obvious to me that the *n*-categorical approach has to be a construction. From that then follows that this construction has to start with $\mathbf{ob}_{\mathbf{AM(0)}}$ since the dimensions of the subsequently constructed $\mathbf{ob}_{\mathbf{AM(0\text{-}i)}}$ (possibly *up to* $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$) will *increase*.[11] Yet this will be a rather uncommon sort of increase, namely an increase - so to say - in negative direction.
But that directedness of the respective *n*-categorical construction doesn't matter at all for the assumed underlying *process* as such. For, if one could imagine such a process as being 'real', then of course such a 'real' process would have to start with an object of a respectively lowest dimension like $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$ or an appropriate $\mathbf{ob}_{\mathbf{AM(0\text{-}i)}}$. This relates obviously to the problem b) mentioned above. But even if it might be hard to envision or imagine such a pre-natural mêontological process one has - at least in the context of this argument - to assume such a process. And therefore questions of attainability of objects like $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$ do not matter in the topos-theoretical perspective. The intrinsic 'dynamics' of the topos **PrePhys** had certainly to start with $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$ (or an appropriate $\mathbf{ob}_{\mathbf{AM(0\text{-}i)}}$) and to terminate in $\mathbf{ob}_{\mathbf{AM(0)}}$.[12]

I've introduced the topos of pre-physical emergence **PrePhys** for mainly two reasons. At first to give a 'consistent' model of the most primordial emergence of structure, namely an even pre-physical emergence. By 'consistent' here I mean 'consistent' with a mode of emergence, which has also applications beyond just that pre-physical emergence itself. That 'mode of how emergence works' can in my view be described by an unfolding of new levels which again in some cases can be modelled by categorification[13], i.e. described by the means of dimensional increase as seen in *n*-category theory. Also foreshadowed in **PrePhys** is the - in my view - very important aspect of a heterotopic mapping[14], i.e. that emergent transitions are essentially characterised by a (stratified) increase of the topological genus, i.e. an emergent higher level has a kind of 'higher dimensional' topological genus, an example for

---

[11] Such a *construction* obviously starts with the object of an at least minimal attainability (which is also the object with the respectively highest dimension), namely $\mathbf{ob}_{\mathbf{AM(0)}}$, i.e. it starts where a pre-natural mêontological 'process' should end. In contrast to $\mathbf{ob}_{\mathbf{AM(0)}}$ the $\mathbf{ob}_{\mathbf{AM(0\text{-}i)}}$ are significantly less attainable and $\mathbf{ob}_{\mathbf{AM(0\text{-}\omega)}}$ is definitely unattainable

[12] Again a bit puzzling is that - to my understanding - such intrinsic dynamics (or the generative structure) of **PrePhys** would by itself show no stratified structure. Thus it seems to me that emergentic features hide in pre-natural mêontological contexts not less than they do in natural ontological ones.

[13] I use this term - which had originally be invented by L.Crane - following J.Baez and J.Dolan. Cf. J.Baez, J.Dolan, Categorification, math.QA/9802029 5 Feb. 1998

[14] The term 'heterotopic mapping' had been suggested to me by Hans van den Berg to describe a 'genus increasing' transformation.

that is 'Antoine's necklace' (a chain the links of which are chains the links of which again are chains etc.). In the case of **PrePhys** attaching the feature of topological genus might be rather metaphorical. Yet it will turn out that together with increasing ontological hardening by the same mechanism which works in **PrePhys** also the feature of topological genus will emerge - soon.

3 The topos of primordial physical emergence (PrimTor)

The following figure FIG.3 shows again a mathematical structure which bears toposic as well as *n*-categorical features, namely the topos of primordial physical emergence which I've called **PrimTor**.

PrimTor is meant to be a topos - so to speak - just on top of **PrePhys**. The objects of **PrimTor** are again *n*-categories but this time there is an important change to note in contrast to the mix up of *n*-categorical and toposic aspects which we have seen in the case of **PrePhys**. In **PrePhys** the virtually infinitary cascade of *n*-categories $\mathbf{ob}_{AM(-i)}$ all had been *strict n*-categories, which are *n*-categories the *objects* of which are respective (*n-1*)-categories themselves. A classical example for such a *strict n*-category is the category of categories **Cat**.

Yet in the case of **PrimTor**, which has just two subobjects (namely $\mathbf{ob}_{AM(0)}$ and $\mathbf{tor}_{AMob(0)}$), the category $\mathbf{tor}_{AMob(0)}$ (which stands for the terminal object) is a *weak n*-categorical extension of the other subobject $\mathbf{ob}_{AM(0)}$ which stands in **PrimTor** for the initial object (but is already known as the terminal object of **PrePhys**). I.e. $\mathbf{tor}_{AMob(0)}$ is a *weak n*-categorical dimensional increase of $\mathbf{ob}_{AM(0)}$ with $\mathbf{am}_{ob(0)}$ (i.e. the automorphism of $\mathbf{ob}_{AM(0)}$ onto itself) as its object and $\mathbf{am}_{amob(0)}$ (i.e. the automorphism of $\mathbf{am}_{ob(0)}$ onto itself) as its *n*-morphism.

This then generates a sort of rolled-up or toroidal plane which would - embedded in a three-dimensional space - look like a cable torus.

Therefore I called this *n*-categorical extension of $\mathbf{ob}_{AM(0)}$ '$\mathbf{tor}_{AMob(0)}$'.

Yet this *weak n*-categorical dimensional increase (resp. 'enlargement', 'enrichment' or 'unfolding') is also of a particular philosophical relevance. Since in contrast to the *strict n*-categorical dimensional increase in **PrePhys** the *weak n*-categorical enrichment in **PrimTor** shows the most defining feature of a real emergent transition, namely a change in the making-up of the intended entities in question. I.e. the succeeding entities in an emergent transition have essentially to differ in their very structure from the preceding ones they stem from. *Weak n*-categorical dimensional enrichment or *categorification* is particularly apt to serve as model for emergent transitions or *complexification* since the *n+1*-category is not an enrichment of the objects of the *n*-category but takes the 'dynamical' aspect of the respective *n*-category as its base.[15] This implies an important 'step aside' which then allows to give a model of the chain of beings not just as a quite linear and continuous sequence of variations but rather of a discontinuous unfolding of an emergent hierarchy of levels. **PrimTor** is meant to represent the most elementary of such levels.

[15] Cf. P.Eisenhardt, D.Kurth, Complexity Categorified; in: Implications, Scientific Aspects of ANPA 22, (ed. Keith Bowden), London 2001

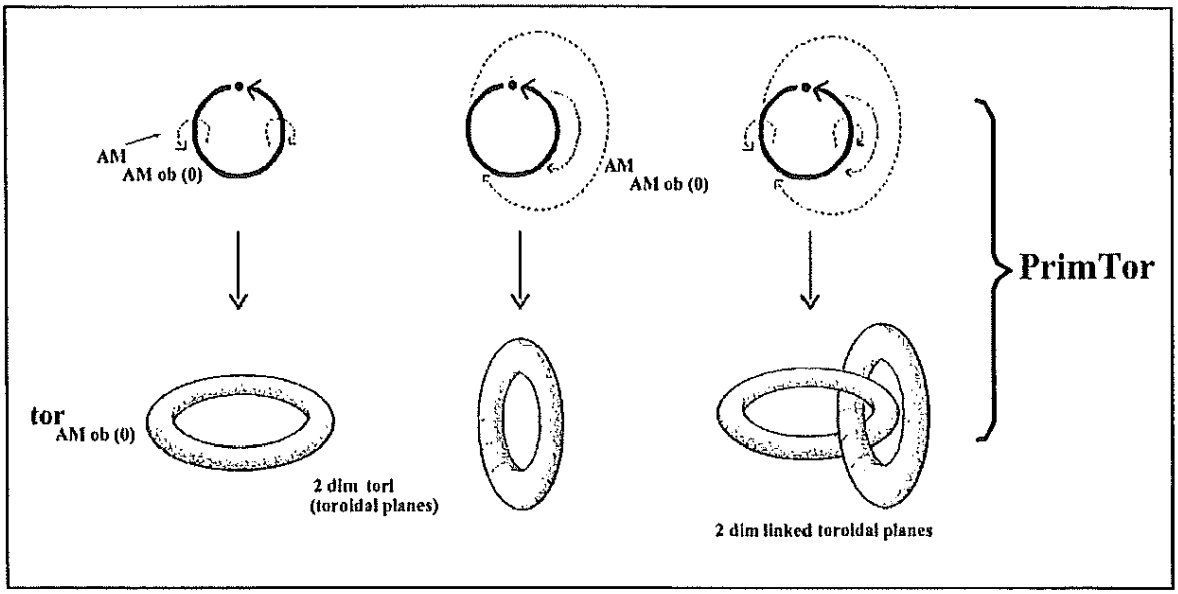


FIG.4: The topos of primordial physical emergence **PrimTor**: a *weak n*-categorical enrichment

For **PrimTor** the following definition holds:

PrimTor: $\{\mathbf{tor}_{\mathbf{AMob(0)}}, \mathbf{ob}_{\mathbf{AM(0)}}\}$

with $\mathbf{ob}_{\mathbf{AM(0)}}$ as initial object and $\mathbf{tor}_{\mathbf{AMob(0)}}$ as terminal object

In the light of the (anti)epistemological distinctions I've mentioned in the introduction to this paper[16] between objects which cannot be intended to be entities (i.e. purely intelligible, formal or mathematical objects) and objects which can be intended to be entities (i.e. potentially physical objects) the tori (or rolled-up toroidal planes) $\mathbf{tor}_{\mathbf{AMob(0)}}$ then are intended to represent such later potentially physical objects as being emerged from an underlying level of purely intelligible or mathematical objects ($\mathbf{ob}_{\mathbf{AM(0)}}$). And such an emergence then would also have to be seen as the emergence of potential physical existence from a underlying state of no physical existence and this again would provide the conceptual means of overcoming the seeming paradox of reductionism - mentioned in the introduction of this paper - without abandoning the reductionist project as such.

3.1 Linked toroidal objects

Since there is no limitation attached for the higher dimensional toroidal objects $\mathbf{tor}_{\mathbf{AMob(0)}}$ concerning either anything like 'orientation' or the number of the n-morphisms $\mathbf{am}_{\mathbf{amob(0)}}$, which essentially constitute $\mathbf{tor}_{\mathbf{AMob(0)}}$, there could very well be various ***linked*** **toroidal objects** come as a result of the respective n-(auto)morphisms $\mathbf{am}_{\mathbf{amob(0)}}$.
Such linked toroidal objects then could be understood as something like elementary chain-elements, which again could be rated as first instances of what I regard to be the quintessential property of - in this case yet very elementary and consequently very 'simple' - 'complex systems', namely topological connectivity.

[16] Cf. also D.Kurth, Actual Existence and Factual Objectivation, in: Movements, Philosophical Aspects of ANPA 23 (Proceedings of ANPA 23), Arleta D. Ford (ed.), London 2002 and D.Kurth, The Tower of Turtles, in: The Philosophical Aspects of ANPA 24, Proceedings of ANPA 24, to appear

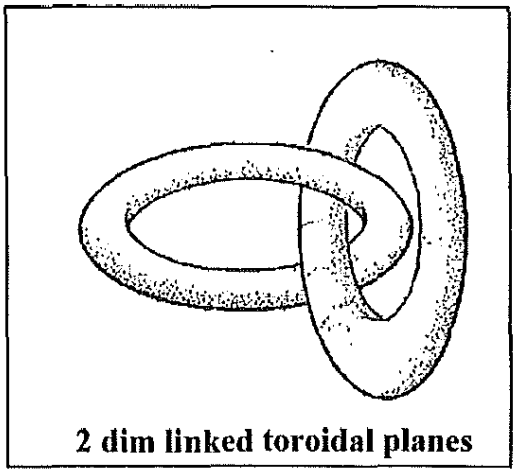


FIG.5: 2 dim linked rolled-up toroidal planes (embedded in $\mathbb{R}^3$): a first instance of potential physical complexity

The philosophical significance of the **linked toroidal objects** shown in FIG.5 lies in my view in the assumption that physical existence is linked to an at least minimal degree of complexity and that - as laid out above - the very feature of complexity is topological connectivity. That leads to the consequence that I doubt it being meaningful to speak of 'elementary physical objects', if by 'elementary' is meant 'simple' or 'of no degree of complexity'. The reason for this point of view is of course that it is consistent with my view of reductionism and the required action to overcome its seeming paradox.[17] Just to make this point as clear as possible: In my view there can't exist an absolutely elementary or simple 'physical entity', i.e. a physical entity of an absolutely least complexity. Quite on the contrary I hold that even the most elementary physical objects stem from mathematical ones and by this they already inherited a respective degree of complexity and since the physical objects emerged out of the mathematical ones their proper rank of complexity is also higher than that of their respective mathematical predecessors. Thus in my view the only objects of a truly least complexity could only be mathematical objects, e.g. of the kind of $\mathbf{ob}_{\mathbf{AM(0-i)}}$.
In the light of these premises I want to highlight the transitorial as well as hermaphrodite nature of **PrimTor**. I.e. **PrimTor** represents just this transitorial stage in the emergence of physical existence, in which its terminal objects $\mathbf{tor}_{\mathbf{AMob(0)}}$ are no more purely mathematical and not yet proper physical ones.[18] And the ***linked* 2-dim rolled-up toroidal planes** then would just be a step further in the direction of emerging physical existence.

[17] Cf. D.Kurth, The Tower of Turtles, in: The Philosophical Aspects of ANPA 24, Proceedings of ANPA 24, to appear

[18] And this provides a good excuse for a surely surprising and perhaps overly precautionary note, namely that an essential part of my argument in this papers relies exclusively on this mentioned double-nature of **PrimTor**. This relates to the emergence of physics out of mathematics. The proclaimed emergence of an elementary (almost) physical state up from a purely mathematical one would precisely take place in the transition from the $\mathbf{ob}_{\mathbf{AM(0)}}$ to the $\mathbf{tor}_{\mathbf{AMob(0)}}$ in **PrimTor**. The infinitary negative ascending hierarchy of $\mathbf{ob}_{\mathbf{AM(-i)}}$ in **PrePhys** is not required for that emergent transition but just for providing a non-paradoxical infinite regress for the reductionist program.

4 Interpretation

4.1 Emergence categorified: *weak 2*-categorical localic/quantal spintori

After coming so near to the advent of physical existence a question carefully avoided till now might eventually come across and can no longer be evaded. This is the question of what particular mathematical objects the $\mathbf{ob}_{AM}$ actually are?
My answer to this question might appear as predominantly clear, but then a bit undecided as well: The 'natural' candidates for being the real $\mathbf{ob}_{AM}$ are either *locales* or *quantales*.
The reasons for this sort of choice (or choices) should be rather obvious. At first the entire context in which the $\mathbf{ob}_{AM}$ undergo their emergent endeavours - namely **PrePhys** and **PrimTor** - is thoroughly toposic. Locales as well as quantales again are of a toposic nature, thus having them as the subobjects of **PrePhys** and **PrimTor** would at least satisfy principles of simplicity as well as selfconsistency.
Yet for the argument in this paper another reason is at least as relevant. Locales (as well as quantales) are just by their very nature ideal candidates for giving the 'Leibniz-point objects'[19] (once envisioned and postulated by Leibniz[20] in his *Theoria motus abstracti*), which had been the starting point of the considerations in this paper, a most rigorous mathematical setting, i.e. a reformulation in the terms of topos theory.

Locales are topos-theoretical doubles of the open sets of set theory and bear also a strong resemblance of the monads of non-standard analysis. All these concepts have in common that they somehow are apt to substitute the ordinary point by a kind of rather infinitesimal element which has - even if not proper parts in the sense of Leibniz then at least - an internal structure.[21] An elementary explanation of the notion of a locale in a somewhat broader context (which has yet the advantage to put this explanation in just that context which has been the starting point of this paper, namely that of topos theory) had been given by Saunders Mac Lane and Ieke Moerdijk:

> "A topos is, in a suitable sense, a generalized space, so should have (generalized!) points. Indeed, at a given point x of an ordinary topological space X, one can erect each set A as a sort of "skyscraper" sheaf A_x concentrated around the point x. The resulting mapping from the category of sets into that of sheaves on X is, in fact, the direct image of a geometric morphism **Sets** $\rightarrow$ Sh (X). But an arbitrary topos $\mathcal{E}$ may not have enough "points" **Sets** $\rightarrow$ $\mathcal{E}$ in this sense. In order to develop an adequate comparison between topoi and spaces, it is useful to alter the definition of a space by describing a space not in terms of its points,

[19] Cf. the introduction to this paper. Cf. also D.Kurth, The Tower of Turtles, in: The Philosophical Aspects of ANPA 24, Proceedings of ANPA 24, to appear

[20] Specifically as subobjects of a topos $\mathbf{PrePhys}_{loc}$ locales $\mathbf{loc}_{AM(-i)}$ might also provide a conceptual basis for reconciling the hitherto seemingly contradictory relationalist and dynamist tendencies in Leibniz' natural philosophy. The relata which are supposed to constitute a certain space then could be seen as respectively lower dimensional (abstract) spaces themselves being constituted by relata which again are respectively lower dimensional abstract spaces related to the one they constitute and themselves again being constituted by another set of such a kind of relata etc.
That possible reconciliation then would be just due to the double nature of $\mathbf{PrePhys}_{loc}$ as a topos and a structure of n-categorical un- or enfolding (see above). The relationist aspect then would be carried by its toposic nature and the dynamist aspect by its n-categorical dowry. I.e. in the toposic perspective $\mathbf{PrePhys}_{loc}$ has all the various locales $\mathbf{loc}_{AM(-i)}$ as ordinary subobjects (and in that sense as proper relata(!)), yet seen as elements of the respective *strict* n-categorical structure (which again happens to be $\mathbf{PrePhys}_{loc}$) any strict n-category $\mathbf{loc}_{AM(-i)}$ has also a strict n-1 category $\mathbf{loc}_{AM(-i-1)}$ as its object. And since $\mathbf{loc}_{AM(-i)}$ would just be produced as the automorphism of $\mathbf{loc}_{AM(-i-1)}$ I would rate this as a genuinely dynamical generation in contrast to a merely static relation.

[21] The mathematical definition and an extensive explanation of locales can be found in: S.Mac Lane, I.Moerdijk, Sheaves in Geometry and Logic: A first Introduction to Topos Theory, New York Berlin 1992

> but in terms of its open sets. The objects so defined by a lattice of open sets are called locales. Since sheaves can be described in terms of coverings by open sets, one can construct a topos Sh(X) consisting of all the sheaves of sets on such a locale X. Moreover, any "continuous" map $Y \rightarrow X$ between locales gives rise to a geometric morphism Sh(Y) $\rightarrow$ Sh(X) between such sheaf topoi."[22]

In this paper locales $\mathbf{loc_{AM}}$, which would stand for the originally introduced $\mathbf{ob_{AM}}$, then would have to be seen as - in an emergence-theoretical sense preceding - constituents of the physical space and not as (in whatever ways derived) parts of it and therefore they would utterly lack the very feature of a pre-particle or of a spatial entity. This is in distinct contrast primarily to loops, which had been designed just as elementary *parts* of physical space, i.e. derived as discrete elements of space based on canonical quantization of the relativistic space-time manifold.[23]

This makes a significant difference because even if one may claim that such an eventually uncovered quantized and thereby discrete deep-structure of physical space would be the underlying and even - in that sense - 'true' structure of physical space one cannot claim that physical space would be emergent from that structure but just the opposite, namely that space always and ever was made up by that elementary structure (and thus *did not emerge* out of a radically non-spatial pre-structure).

Yet to claim that space itself is radically emergent from a pre-spatial structure requires a completely different approach, since - if by the epithet 'spatial' here a *physical space* is suggested - such a pre-spatial structure would have to be of an essentially non-spatial making.

Locales as well as their non-commutative siblings, quantales[24], now are just such mathematical objects, which neither had been derived from any concepts of physical space nor designed for the purpose to provide elements for a physical space to be built up by (or on) them. I.e. locales (and quantales) must *not* be seen as - in what ever way diluted or diminihed - *parts* of physical space, not at least because they cannot be taken as physical entities at all.

But if the assumed most imperceptibly emerging physicality does not stem from the *objects* in the topos(es) $\mathbf{SpinTor_{loc}}$ (resp. $\mathbf{SpinTor_{qtl}}$) - (both) shown (in a sort of superposition) in FIG.6 (which are just the respectively interpreted versions of **PrimTor,** see FIG.4) then - since nothing else is left - any possibly emerging physicality must stem from the respective weak n-morphisms in $\mathbf{SpinTor_{loc}}$ (resp. $\mathbf{SpinTor_{qtl}}$), i.e. the $n+1$ (auto)morphisms $\mathbf{sp_{AMloc(0)}}$ (resp. $\mathbf{sp_{AMqtl(0)}}$) of the n-(auto)morphism $\mathbf{AM_{loc(0)}}$ (resp. $\mathbf{AM_{qtl(0)}}$) onto themselves.

---

[22] S.Mac Lane, I.Moerdijk, Sheaves in Geometry and Logic: A first Introduction to Topos Theory, p 8

[23] In such a sense (as well in many others aspects) the role attached in this paper to locales (or quantales) differs decisively from e.g. the role of Wilson Loops in the loop-theoretical approach to Quantum Gravity.

[24] Most of what has been said about locales above also relates to quantales, which are in fact a sort of quantized locales, i.e. locales which satisfy an additional noncommutative rule. Quantales had been originally introduced by C.J..Mulvey in 1986. For a detailed introduction into the subject cf. C.J.Mulvey, M.Nawaz, Quantales: Quantal sets, in: Non-classical Logics and their Application to Fuzzy Subsets: A Handbook of the Mathematical Foundations of Fuzzy Set Theory, Kluwer 1995, pp 159-217; somewhat more related to the topic of this paper is: C.J.Mulvey, J.W.Pelletier; On the Quantisation of Points, J.Pure Appl. Algebra 159 (2001) 231-295

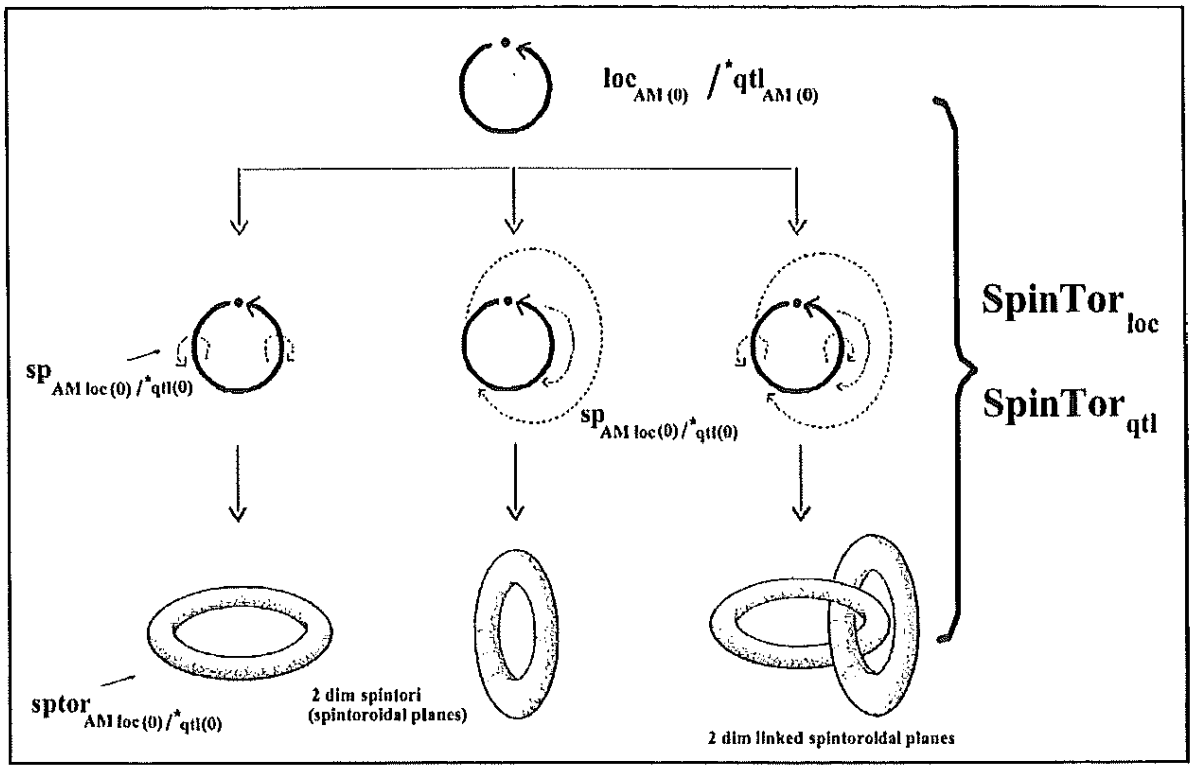


FIG.6: The topos(es) of emergence: SpinTor loc /*qtl

* [The dash (/) in FIG.6 signifies that the objects in question are always supposed either to be read as proper *locales* or as their non commutative siblings, namely *quantales*.]

For $\mathbf{SpinTor}_{loc}$ the following definition holds:

$\mathbf{SpinTor}_{loc}$: $\{\mathbf{loc}_{AM(0)}, \mathbf{sptor}_{AMloc(0)}\}$

with $\mathbf{loc}_{AM(0)}$ as initial object and $\mathbf{sptor}_{AMloc(0)}$ as terminal object

Respectively the following definition holds for $\mathbf{SpinTor}_{qtl}$:

$\mathbf{SpinTor}_{qtl}$: $\{\mathbf{qtl}_{AM(0)}, \mathbf{sptor}_{AMqtl(0)}\}$

with $\mathbf{qtl}_{AM(0)}$ as initial object and $\mathbf{sptor}_{AMqtl(0)}$ as terminal object

Thus the emergence of potential physicality is - in my view - utterly based on the dynamical - which here are in fact *the same as the emergentic* - aspects of the mathematical structure(s) ($\mathbf{SpinTor}_{loc}$ resp. $\mathbf{SpinTor}_{qtl}$) in question, namely the aspects, which are related to the *weak n*-categorical entrails of the topos(es) $\mathbf{SpinTor}_{loc}$ (resp. $\mathbf{SpinTor}_{qtl}$, i.e. the *weak 2*-categorical (auto)morphisms $\mathbf{sp}_{AMloc(0)}$ (resp. $\mathbf{sp}_{AMqtl(0)}$).[25] And this is perfectly in line with the philosophical premises which led me to all these conclusions[26] and which originated in Leibniz' radical dynamism as displayed in his *Theoria motus abstracti* (as well as in many others of his writings).

[25] But in sharp contrast to the similar situation in the case of **PrePhys** here the mentioned "*weak n*-categorical entrails" are no more and not at all completely independent of the toposic structure of $\mathbf{SpinTor}_{loc}$ (resp. $\mathbf{SpinTor}_{qtl}$), but - just to the contrary - in the case of $\mathbf{SpinTor}_{loc}$ (resp. $\mathbf{SpinTor}_{qtl}$) the (toposic) morphisms, which lead from the initial object(s) $\mathbf{loc}_{AM(0)}$ (resp. $\mathbf{qtl}_{AM(0)}$) to the terminal object(s) $\mathbf{sptor}_{AMloc(0)}$ (resp. $\mathbf{sptor}_{AMqtl(0)}$) are just *identical* with the *weak n*-categorical (auto)morphism(s) $\mathbf{sp}_{AMloc(0)}$ (resp. $\mathbf{sp}_{AMqtl(0)}$).

[26] Cf. D.Kurth, The Tower of Turtles, in: The Philosophical Aspects of ANPA 24, Proceedings of ANPA 24, to appear; and D.Kurth, Actual Existence and Factual Objectivation, in: Movements, Philosophical Aspects of ANPA 23 (Proceedings of ANPA 23), Arleta D. Ford (ed.), London 2002

But then what could that insinuated physical feature of $\mathbf{sp_{AMloc(0)}}$ (resp. $\mathbf{sp_{AMqtl(0)}}$) ever be? Since we haven't been parsimonious with exacting suggestions to the imaginative benevolence and faculty of the readers, we feel entitled to stress at least that benevolence (of the remaining ones) a little further.

I take or interpret the $\mathbf{sp_{AMloc(0)}}$ (resp. $\mathbf{sp_{AMqtl(0)}}$) as a sort of (physical) spin. And I would even extend this interpretation to the preceding automorphism(s) $\mathbf{AM_{loc(0)}}$ (resp. $\mathbf{Am_{qtl(0)}}$), i.e. I would take them as being spins of either locales or quantales. With this surely exacting suggestion we now have reached the decisive point of the whole argument, which I put forward in this paper. The only excuse for all these exactions is the fact that the notion of 'emergence' inherently has a tinge of unreasonableness, since it implies to find a way (a sort of continuity) across an abyss (of a radic al discontiuous nature).

As always in the theory of emergence also in the case of the relation of the concepts of 'automorphism' on the one hand and of 'spin' on the other the reductionist view works fine and the proper emergentist one much less fine, if at all. This becomes obvious when one agrees that the physical spin can - as an abstract symmetric rotation - nicely be modelled as an instance of an automorphism, yet that an automorphism can obviously not be seen as an instance of a spin, simply because it lacks the physical features of a spin.

Yet as always in the theory of emergence one has to overcome the undeniable reasonableness of this objection. Since by not doing so one has to pay a much higher price in terms of reasonableness, i.e. one has either to admit the existence of miracles, e.g. in the case of the emergence of life up from prebiotic conditions, or of potential paradoxes as in the case of the reductionist project with respect to the question of the ontological status of a first physical entity.

So I hold that $\mathbf{SpinTor_{loc}}$ (resp. $\mathbf{SpinTor_{qtl}}$) provides a model of how automorphisms of the sort of $\mathbf{sp_{AMloc(0)}}$ (resp. $\mathbf{sp_{AMqtl(0)}}$) and their 'predecessors' $\mathbf{AM_{loc(0)}}$ (resp. $\mathbf{Am_{qtl(0)}}$) once could have find a way across the abyss between the mêontology of mathematical objects and the ontology of intended physical entities by metamorphosing into spins.

4.2 From spinchains to contiguity: emergent topological connectivity = complexity

Yet of course we have still left out some further exacting details of $\mathbf{SpinTor_{loc}}$ (resp. $\mathbf{SpinTor_{qtl}}$), namely the status of their terminal objects $\mathbf{sptor_{AMloc(0)}}$ (resp. $\mathbf{sptor_{AMqtl(0)}}$).

Now, these are apparently static images of the automorphisms $\mathbf{sp_{AMloc(0)}}$ (resp. $\mathbf{sp_{AMqtl(0)}}$) of which they are 'made of'. I.e. insofar as they are statified automorphisms they are ontified as to be a sort of elements. And they are supposed to be elements - quite literally. This becomes evident in the bottom line of FIG.6 in the case of the object called '2-dim linked spintoroidal planes'.

I.e. the real potential physical role of $\mathbf{sptor_{AMloc(0)}}$ (resp. $\mathbf{sptor_{AMqtl(0)}}$) as elements lies exclusively in their potential role as elements of a spinchain, i.e. a spinchain just of such elements as $\mathbf{sptor_{AMloc(0)}}$ (resp. $\mathbf{sptor_{AMqtl(0)}}$).

Such spinchains then might - appropriately increased - grow into more complex physical entities. A possible example could be entities like 'strings', the proposed elements of such entities, namely $\mathbf{sptor_{AMloc(0)}}$ (resp. $\mathbf{sptor_{AMqtl(0)}}$) then being a sort of 'parton' of such spinchains or 'strings'. Yet serious physical stuff was never the topic of this paper so we stick to our philosophical suggestions. As a somewhat enlarged illustration of these philosophical suggestions see FIG.7.

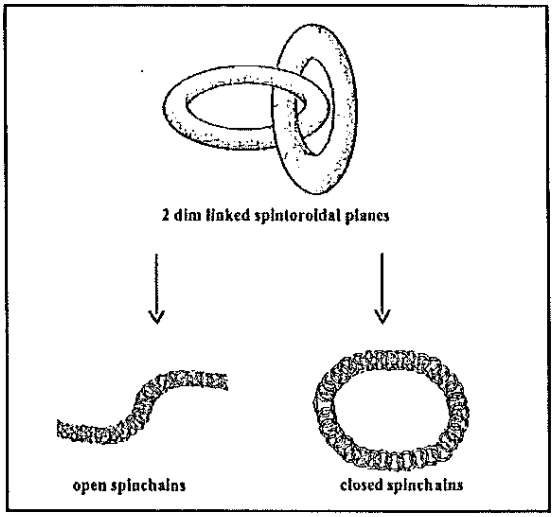


FIG.7: spinchains: emergent topological connectivity built-up of linked spintori

In this respect what matters most with the spinchains shown in FIG.7 is that the emerging physicality - claimed to be illustrated there - is essentially due to the growing topological connectivity which is as well characterised by the contiguous structure (of a chain).

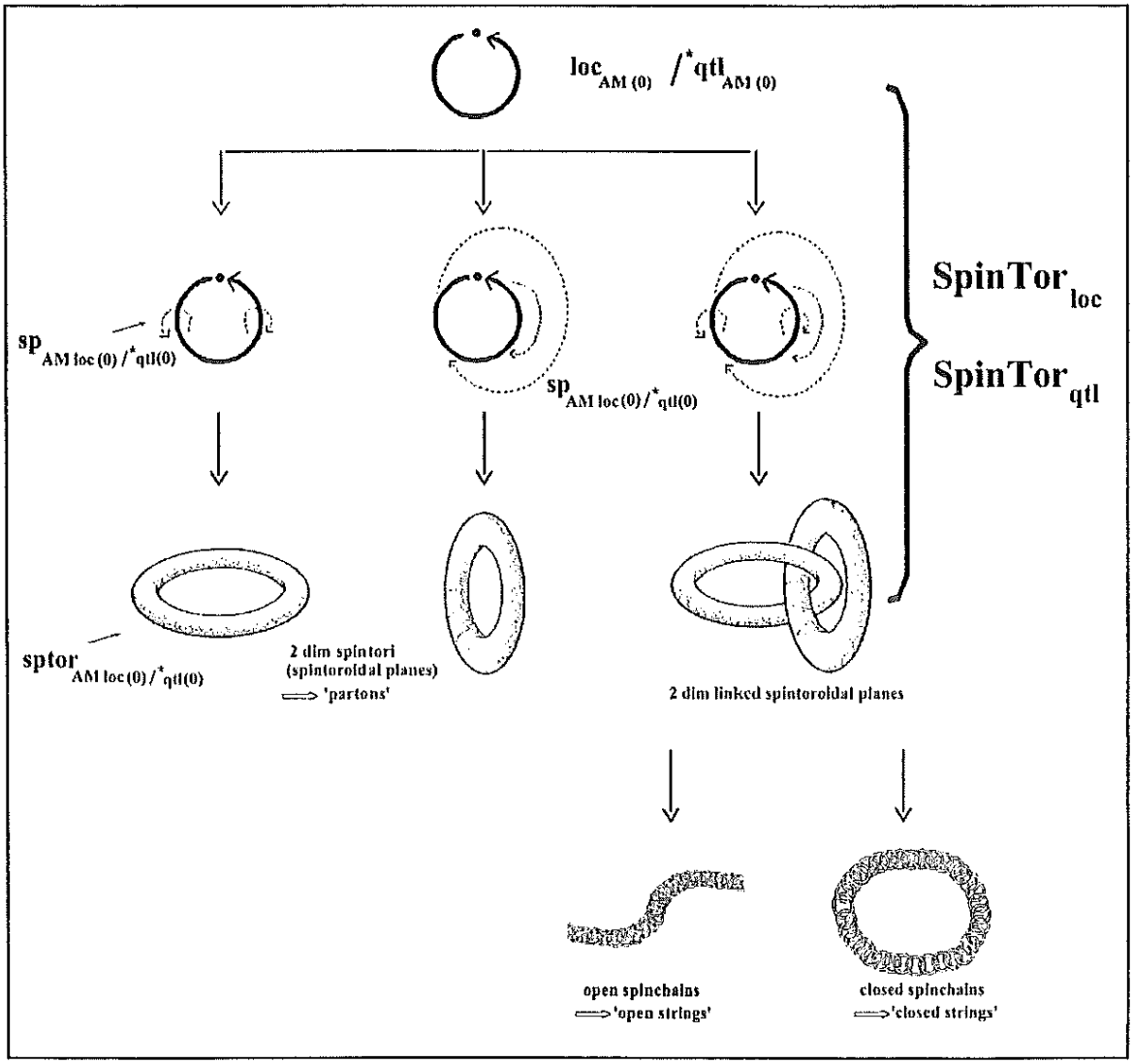


FIG.8: The topos(es) of emergence. And elementary spinchains (made of linked spintori) as first instances of emergent topological connectivity.

As the mentioned growing topological connectivity is producing more of that contiguous structure(s) by its own further growth it also produces more and more of potential actual physicality. Thus - I hold - that exactly such topological connectivity or contiguity is the underlying structure (and measure) of physical complexity and even the most elementary physical entities already are characterised by an increased complexity related to that of the mathematical objects from which they stem.[27]

Since this paper was - for good reason - not about physics at all but just about to think of it as emerging, it now already overstretched its scope. Therefore it may now come to an end after having presented you instead of a summary an overview of the proposed topos(es) of emergence in FIG.8.

[27] This is again in line with my claim that any assumed most elementary physical entity already has to have a certain degree of complexity which directly stems from (the increase of complexity having taken place in the pre-physical emergence of) its mathematical predecessors.

Alternative Natural Philosophy Association

Statement of Purpose

1. The primary purpose of the Association is to consider coherent models based on a minimal number of assumptions, so as to bring together major areas of thought and experience within a natural philosophy alternative to the prevailing scientific attitude. The Combinatorial Hierarchy, as such a model, will form an initial focus of our discussions.

2. This purpose will be pursued by research, publications and any other appropriate means including the foundation of subsidiary organisations and the support of individuals and groups with the same objective.

3. The Association will remain open to new ideas and modes of action, however suggested, which might serve the primary purpose.

4. The Association will seek ways to use its knowledge and facilities for the benefit of humanity and will try to prevent such knowledge and facilities being used to the detriment of humanity.

Organisation (altered to reflect current situation by KGB)

1. The Executive Council is the governing body of the Association. It consists of:

   (a) All past Presidents of the Association.

   (b) Officers (President, Treasurer and Co-ordinator).

   (c) Ordinary members nominated by classes (a) and (b), who serve for three years, with the possibility of re-nomination.

2. Members of the Association are (a) members of the Executive Council and (b) others nominated by the members and approved by the Executive Council.

3. The President is the official representative of the Association in external affairs, and has the responsibility for calling meetings of the Membership, at least annually, for the determination of overall policy.

4. The Treasurer is the responsible financial officer of the Association for the receipt and disbursement of funds.

5. The President and Co-ordinator may be paid an appropriate salary for their services, funds permitting. These services will include the organisation of meetings and the editing of the Proceedings of such meetings for publication, co-ordination of and participation in the research activities of the Association, preparation when appropriate of research reports and publication of such reports, and other such duties as may be assigned.

6. President, Co-ordinator and Treasurer may, as appropriate, receive funds for travel expenses, secretarial help, etc.

7. The Executive Council has selected an independent Advisory Board. It may adopt its own rules for the operation and replacement of members. The Executive Council may nominate candidates to the Board. Any member of the Board, or the Board collectively, may make recommendations to the Executive Council, or directly to the Membership. Action taken on such recommendations must be promptly reported by the Executive Council to the Board in writing.

8. Dues are currently £20.00 per annum.

Executive Council: Dr. John Amson, Dr. Ted Bastin, Mr. Anthony M. Deakin (Chairman), Dr. Tom Etter, Ms. Arleta Griffor, Prof. Louis Kauffman, Prof. Clive W. Kilmister, Dr. Michael Manthey, Prof. H. Pierre Noyes, Dr. David Roscoe, Dr. Fredric S. Young, Prof. Rainer Zimmerman.

President: Dr. Keith Bowden, 139 Sandringham Road, Barking, Essex, IG11 9AH, UK.

[Tel: 0208 594 5064, Email: k.bowden@physics.bbk.ac.uk].

Treasurer: David Roscoe, Department of Applied Mathematics, Sheffield University, Sheffield S3 7RH.

[Tel: 01142 235 2785, Email: d.roscoe@sheffield.ac.uk]

Co-ordinator: Arleta Griffor, 37a Shepherds Way, Rickmansworth WD3 7NN.

[Tel: 01923 461679, Email: a.griffor@physics.bbk.ac.uk]

Advisory Board: M. Horner (Chairman), Profs. G.F. Chew (Berkeley), C. Isham (Imperial College), M. Redhead (Cambridge), N. Cartwright (LSE), C. W. Kilmister (London, retired), H. Pierre Noyes (Stanford, retired).