Aspects II

Proceedings of ANPA 20

K. G. Bowden, *Editor*

Proceedings of the 20th Annual International Meeting of the

Alternative Natural Philosophy Association

Wesley House, Jesus Lane, Cambridge,
September 1998

published by **ANPA**
c/o Dr. K. Bowden,
Theoretical Physics Research Unit,
Birkbeck College, Malet St, London WC1E 7HX.

May 1999

Aspects II: Proceedings of ANPA 20/Keith G. Bowden, *Editor*

ISBN 0 9526215 4 1

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

Contents

ANPA20 Editorial

Keith Bowden

My original intention after the nonappearance of ANPA19 was to publish an ANPA19/20 double issue this year. However, when all the copy for both issues was in, it became clear that there were well over 500 pages and that the new binding procedures at the University of London would not allow for this in a single volume. **Aspects** (in honour of Clive Kilmister's recent work) thus became split into **Aspects I** and **Aspects II** retaining, I hoped, some of the flavour of a double issue. In particular this issue contains Arleta Griffor's **"On the Non-Commutative Combinatorial Hierarchy"**, which builds directly on Clive's work in the companion volume. I have also included a reprint of an old paper of mine with Clive Kilmister on the CH, originally presented at ANPA West in Stanford, "An extension to Eddington's Fundamental Theory" which a number of people have asked to see. An Appendix of the Proceedings of the two day extended session to discuss the paranormal is also available as **Paranormal Aspects**. I hope everyone likes the new wrap around, perfect bound format, (and I hope the glue holds better than before!) My thanks are due to Patrick and David at the Print Unit in SOAS for producing these volumes (and to Christine at Birkbeck Print Unit for the Appendix).

The last editorial I wrote (other than that in the Newsletter, from which this is partially cribbed) was the one in the Proceedings of ANPA18, so there is a fair amount of news since then. I left UEL (in disgust) and am now in the Theoretical Physics Research Unit at Birkbeck College. TPRU is what is left of David Bohm and Basil Hiley's Department of Physics after Baroness Blackstone - one time Master of Birkbeck - closed it down. Its future remains in the balance, but is currently assured to September 2001. The atmosphere at Birkbeck is good and productive and a number of TPRU members are also ANPA members, in particular Arleta Griffor and Owen Maroney. Saral Bohm and Don and Anna Factor attended ANPA20 and helped organise a successful Bohm Dialog session. Our thanks to them. Basil Hiley has kindly agreed to be our Special Guest Speaker for ANP21. I am hoping that he will talk on the Algebra of Process. I have included a recent interview of Basil by Arleta Griffor, which should set a good background for this talk.

There has been some rearrangement within ANPA. Pierre Noyes and Clive Kilmister have both resigned from the Executive Council and Viv Pope has resigned as Secretary and Newsletter Editor. Our heartfelt thanks to them for all the work they have done in the past. Mary Pope has generously agreed to act as Secretary for the coming year only in order to arrange the ANPA21 Conference, then we will need a new Secretary, without whom there will not be an ANPA22. All volunteers and nominations to me as soon as possible please. There may be the possibility of small remuneration for carrying out this task. Arleta Griffor is now Newsletter Editor and she and David Roscoe are now on the Executive Council. Nancy Cartwright from LSE and Clive Kilmister have accepted places on the Advisory Board. It is also worth mentioning a couple of moves. Ted Bastin now lives not too far from Viv Pope in South Wales. Ted has been in hospital recently (he returned to Cambridge for this) but seems now to have made an excellent recovery. Viv, too, has been in hospital recently and we wish him a speedy recuperation. Mike Manthey has moved from Denmark to Colorado. Hope the meditation goes well, Mike.

Peter Marcer and I found ourselves in Helsinki in November 1998 for the European Pathfinder Conference on Quantum Computation. All in all I found this a useful visit, although it was somewhat marred for some by the sudden rush when the man from the EC with the chequebook arrived. For those of you who are interested in Peter's views of the week I am sure he would be happy to elucidate. Peter was a founder member of the Pathfinder Group and should be congratulated on helping organise this international event on such an important subject. As a result of the initiative I am looking at producing a bibliography of Quantum Computation with Owen Maroney for Springer as part of the Pathfinder project. Another tentative project is a book on Quantum Theory with David Robson also at TPRU.

Viv Pope's book on Action at a Distance is well underway. I checked out one of the submitted papers on the Los Alamos archive (on structural stability in theory space) and was impressed at the very high standard. It is hard to produce good stuff on this subject. There was an enormous response to the Call for Papers so it will be about twice as long as originally scheduled. It is due to be published later this year under the title 'Action at a Distance - Pro and Contra', edited by A. Chubykalo, N.V. Pope and R. Smirnov-Rueda, published by Nova Science Publishers, Inc., New York. It will contain contributions from Ted Bastin and Clive Kilmister, Franco Selleri, Tom Phipps (a friend of Pierre Noyes at Imperial College), Peter Rowlands, who joined ANPA last year, David Roscoe and Viv Pope, so ANPA members are well-represented. Other contributors include Neal and Peter Graneau, A.K. Assis, E. Comay, G. Galeczki, and C.I. Mocanu. There are 36 contributions altogether, a compilation of physics, mathematics and philosophy which should make for interesting reading.

I am trying to build an archive of older ANPA material with a view to (re)publishing some of it. In particular I am hoping to produce a "Best of ANPA West" and a "Best of ANPA Proceedings". If anyone can throw light on what happened at the earlier meetings and if any material remains I would much appreciate seeing it. I also hope to publish a catalog for ANPA West, with the contents of the Instant Proceedings. Finally, I have mentioned to some members of ANPA that I would like to publish a Bibliography of Combinatorial Hierarchy/ANPA related papers. I would much appreciate it if those who could make a personal bibliography available to me would do so (preferably electronically).

The International Journal of General Systems Special (ANPA) Issue on Physical Systems is finally out with papers by Bastin and Kilmister, Manthey, Bowden, Kauffman, Amson, Parker-Rhodes (Frederick, posthumously) etc. I am in general pleased with the issue which runs to some 250 pages, even with some colour plates, possibly the largest ever IJGS issue, but it is sad that Tom Etter's wonderful paper Process, System, Causality and Quantum Mechanics was withdrawn by the Managing Editor at the last minute. The issue feels somewhat incomplete without this, although I have a scheme to reprint as a book including Tom's paper. Meanwhile it is available on the Los Alamos eprint archive and as SLAC PUB-7890, carefully edited by Pierre Noyes. The Special Issue was unfortunately retypeset by the printers ***after*** we had spent some considerable time proofreading it, and they inevitably managed to introduce a number of mistakes, not the least of which was reporting that John Amson sleeps "on the floor" of his house in Anstruther, rather than "on the third floor"! An errata appears below.

Errata for Special Issue of IJGS

The Special Issue appeared as Vol. 27, Nos. 1-3 (1998) of the International Journal of General Systems.

These are the mistakes I know of up to now. Could you please let me know of any further errors or typos as IJGS wish to publish a proper errata.

1. My affiliation should be as below, and the editor is shown as G. J. Klir who is actually the series editor☹.

2. p.2 para.1 line 15 sleeps on floor -> sleeps on third floor
   p.2 para.1 line 16 Atlantic Ocean -> North Sea

3. p.81 Abstract line 1 structured bets -> structured sets
   p.90 para.4 line 1 set of 162 × 2 -> set of 16 × 2

4. p.28 line 6 denvation -> derivation
   p.28 lines 13-15 should read

$$E=X^{tt}-X^{t}\times H \text{ and } H=-\nabla\times X^{t}/2$$

and

$$\partial H/\partial t=\mathrm{d}H/\mathrm{d}t-(\partial X_j/\partial t)\partial H/\partial X_j$$

5. p.33 Ted Bastin's address is given correctly below the main title, but a corrupted form of my old affiliation is given as his corresponding address at the bottom of the page. This looks as though it was dictated over the telephone by someone with a cold to a non English speaking typesetter in the Phillipines. It should simply be omitted.

6. There are a number of references to Tom Etter's paper Process, System, Causality and Quantum Mechanics (subtitled A Psychoanalysis of Animal Faith) which should have appeared in this issue. It actually appears as SLAC-PUB-7890, August 1998 co-authored with (actually very carefully edited by) Pierre Noyes. It also appears at the Los Alamos eprint archive (http:/xxx.lanl.gov) as quant-ph 9808011.

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 was approved at the last meeting of 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 should be the common language of Physics as usually understood by Physicists. Any other use of language should be carefully explained at the start of the paper and all appropriate definitions included there.

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 papers in any recent issue of the Proceedings. At least, Times Roman, 12 point, ***single sided, two copies,*** is preferred, although I will still accept typescripts in Courier. **10 point is TOO SMALL to be reduced to A5; 14 point is better for short papers.** Main heading capitalised, centred, other headings capitalised to the left. No underlining. At least a one inch bottom margin for footers; page numbers NOT bottom right. *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*. Please help me by conforming to all this as closely as you can. THIS PLEA HAS LARGELY BEEN IGNORED IN THE PAST. PLEASE HELP BY READING IT CAREFULLY. As January 1st has proved impractical in the past the copy date for the Proceedings of ANPA21 is 1st April 2000 (often used as an academic deadline as any fool should know!) 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
0181-594-5064
FOR EDITORIAL ADDRESS SEE LAST PARAGRAPH

False conjectures about cycles

C W Kilmister

Red Tiles Cottage

High Street

Barcombe

Lewes, BN8 5DH, UK

I report briefly on part of the work that John Amson and I have been doing. He will explain his side later.

Early on the Combinatorial Hierarchy was considered in Frederick Parker-Rhodes's original notations, vectors v over the field $\mathbb{Z}_2$ and matrices A operating on them. Part of our effort then to understand this curious construction was to try to put it into some context. An important part was played by the equation $Av = v$ in Frederick's construction. We thought a context for this might be the consideration of the more general situation, $Av = v_1$, $Av_1 = v_2$ and so on. There is only a finite number of vectors v_r and A is non-singular, so eventually the sequence v, v_1, v_2, ... gets back to v. We called a sequence v, v_1, v_2, v, ... a 3-cycle, so that Frederick's original eigenvectors were 1-cycles.

Originally we paid attention to possible cycle-lengths for all n x n matrices. My colleague Albrecht Fröhlich easily proved for us that there will always be a maximal cycle of all the vectors, so of length $n^* = 2^n - 1$. This is because the vectors can be seen as elements of a field of characteristic 2, and there will be left-multipliers u in the field, such that $Av = uv$ with the required property. It is obvious that the cycles for $n = r - 1$ will also be present for $n = r$ (a sub-space construction) so, looking at $n = 3$ in particular, 7, 3 and 1 cycles are possible. To get farther, I should explain a more economical notation for vectors and matrices. I use p (where p is an integer) to denote the vector with 1 in the p th place and zeros elsewhere. Then the vector p + q can be shortened to pq, so long as p and q do not exceed 9. A matrix can be expressed by its columns, so that $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ will be written (23, 3, 1). The usefulness of this notation is that, for instance, (a, b, c)1 = a, (a, b, c)2 = b and so on.

Consideration of (2, 1, 3) shows that a 2-cycle is possible for n = 3 and then (2, 1, 13) exhibits 3 - 13 - 123 - 23 - 3, a 4-cycle. On the other hand, 5- and 6-cycles are impossible for n = 3. For, firstly, a 5-cycle would leave over 2 other elements. These cannot be eigenvectors since any two such add to a third, and they cannot be a 2-cycle since the sum of the elements of any 2-cycle is an eigenvector, again giving three elements outside the 5-cycle. Similarly a 6-cycle, say a - b - c - d - e - f, leaves one other element which must be an eigenvector. But in general a 6-cycle gives rise to a 2-cycle: ace - bdf and to a 3-cycle ad - be - cf. These associated sycles must degenerate into the same eigenvector so that

ace = bdf, ad = be = cf.

Thus abcdef = 0 and also abde = 0. Hence cf = 0, so that a = d, b = e and c = f. The original 6-cycle is not a proper 6-cycle at all. The spectrum of cycle-lengths for n = 3 is therefore 1, 2, 3, 4, 7.

Here are a few general results, much weaker than Fröhlich's. Introduce the belonging relation B:- Br, n means that there is an n x n matrix and an n-vector generating an r-cycle. Thus, Br, 3 if r = 1, 2, 3, 4, 7. Fröhlich's result is Bn*, n and the subspace construction gives

Br, n $\rightarrow$ Br, n + 1,

with the obvious corollary

Br, n $\rightarrow$ Br, n + s, (s = 1, 2, 3...).

Moreover, from the matrix A = (2, 3, 4, ... n, 1) it follows that Bn, n and using this with subspaces gives Bn, n + s. Putting m for n + s so that m - s = n, one gets Bm - s, m. Thus the spectrum for n is full from the n-cycles downwards. Moreover using Fröhlich's result with this, Bn*, n + s or B(m - s)*, m for s = 1, 2, ... One further result is B2n, n + 1 which one can prove by starting with A = (2, 3, ...n, 1) and converting it into the (n + 1) x (n + 1) matrix B = (2, 3, ...n, 1n', n') where n' is the successor of n. From these results one can see at once that Br, 4 if r = 1, 2, 3, 4, 6, 7, 15. Also B5, 4 by A = (2, 3, 14, 34). The spectrum from 8 to 14 is empty, the proof for 13 and 14 being as above for 5 and 6 in three dimensions, but I have only a very clumsy proof of the others.

Back in the sixties the spectrums for n = 3, 4 suggested the fanciful notion that the possible cycle-lengths were related to the spectrum of particle masses known at the time. It seemed worth computing mechanically up to n = 16. Fig 1 shows the correct results for 3 and 4 and the computed ones for 5 and 6. The computation was rather inaccurate, depending on an insufficiently general canonical form of A, and now needs re-doing. Still, the hypothesis seems pretty conclusively falsified.

This year John and I have spent a long time on another mathematical conjecture. Given a matrix A, say for n = 4, producing an eigenvector, a 2-cycle and two 4-cycles, we call $1.2.4^2$ its "cycle-complexion". We wondered if the cycle-complexion determined the matrix. Well, obviously not, because $(PA^rP^{-1})(Pv) = PA^rv$ so that PAP^{-1} has the same cycle-complexion as A for any non-singular P. But does the cycle-complexion determine the matrix up to such similarity transformations? I think it does for n = 2, 3 and for quite a lot of other matrices but we found it hard to prove. And the reason, of course, was because it was false. A counter-example is A = (2, 3, 4, 123), B = (2, 3, 4, 134). A has zero trace but B does not, so they cannot be similar.. Yet A generates:

1 - 2 - 3 - 4 - 123 - 234 - 124, 12 - 23 - 34 - 1234 - 14 - 13 - 24

with eigenvector 134, so has complexion 1.7^2, and B will be found to have the same complexion.

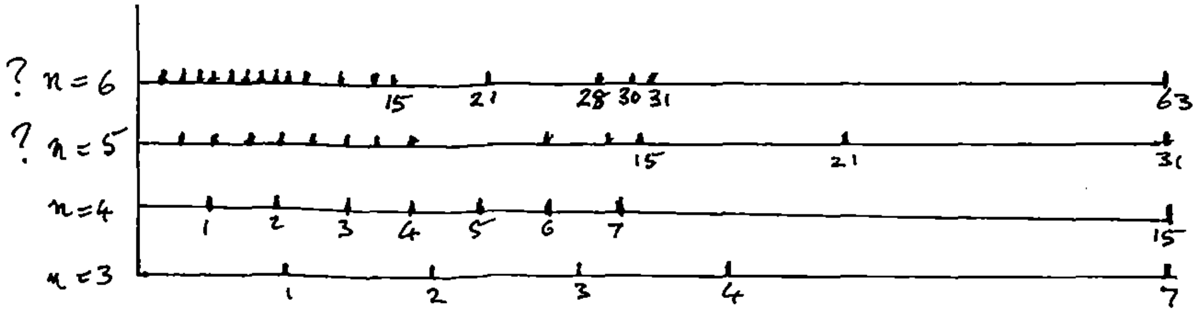


Fig. 1. Cycle-lengths for n = 3 - 6
with scales to make the maximal cycles
agree.

(The 4 and 21 in the upper halves for the odd n are of some interest.)

ON THE NON-COMMUTATIVE COMBINATORIAL HIERARCHY

Arleta Griffor

Theoretical Physics Research Unit
Birkbeck College, University of London
Malet Street, London WC1E 7HX
a.griffor@physics.bbk.ac.uk

1. ON DIFFERENCES AND SIMILARITIES

My contention is that the Process which gives rise to the Combinatorial Hierarchy (CH) can be regarded as *an ordered coming to being of new orders.* The concept of order is understood here in the way it has been introduced and developed by D. Bohm.[1]

According to D. Bohm, the most general way of understanding order is in terms of similar differences and different similarities. The most simple order is that of similar differences, that is, differences that are related by being similar. More complex orders can be understood by considering similarities and differences of *orders* themselves. In this way what is being considered is an order of orders, or an ordered series of orders that is generated (in some ordered way) by similar differences.

In those terms, the Combinatorial Hierarchy can be understood as an order of orders generated by a certain set of similar differences (originating in the Process). The Process, by means of discrimination, distinguishes between similarities and differences. In other words, it distinguishes whether it came upon a *similarity* or a *difference*, and it responds accordingly. The response is again in terms of *similarities* or *differences.*

In this view, *nothing is generated in the Process but similarities and differences.* Labels, (denoted by d_i, d_k), are differences, and discriminations of labels, (denoted by $d_i d_k$), are differences or, indeed, similarities. The Process is thus a generative Process in which differences, similarities, and in the next step, more complex orders of similar differences and different similarities (i.e., 'sets closed under discrimination') etc. arise.

This way of looking at the Process differs from the way it was originally introduced.[2] The difference is not only in renaming *labels* as *differences* and *signals* as *similarities,* but also in assuming that *both* differences and similarities are generated in the Process.

2. The First Level of NCH

2.1 Algebraic structure of $G_{22}^{(0)}$

The first level, $G_{22}^{(0)}$, of the non-commutative combinatorial hierarchy (NCH) is generated by two differences, d_1 and d_2, with an operation of discrimination. The following assumptions determine the structure of $G_{22}^{(0)}$:

1. $d_i d_i = (d_1 d_2)(d_1 d_2) = e_o$ for $i \in \{1, 2\}$
2. $(d_1 d_2)(d_2 d_1) = (e_o)^2 = e$
3. $xe = ex = x$ for any $x \in G_{22}^{(0)}$
4. $x(xy) = (xx)y$
$y(xx) = (yx)x$ for any $x, y \in G_{22}^{(0)}$

The assumptions 1 - 3 are essentially those proposed by C.W. Kilmister.[3] The assumption 4 (of weak associativity) is a *new* assumption which replaces Conway's method of generating the elements of $G_{22}^{(0)}$.[4]

The elements e_0 and e, are two (different) similarities of $G_{22}^{(0)}$. The relationship of similarities and differences of $G_{22}^{(0)}$ is specified by the assumptions 1 - 3.

From **1** - **4** we can prove the following consequences:

C1. $e_0 d_i = d_i e_0$
Proof: from 4, for $x = y = d_i$, we have $(d_i d_i)d_i = d_i(d_i d_i)$, therefore from 1 we have $e_0 d_i = d_i e_0$ ■

C2. $e_0 d_{12} = d_{12} e_0$
Proof: the same as C1, if we take $x = y = d_{12}$ (where d_{12} denotes $d_1 d_2$)[5]

C3. $d_{ik} = e_0 d_{ki}$
Proof: $x = d_{12}$ and $y = d_{21}$, from 4 and 1, we have
$d_{12}(d_{12}d_{21}) = e_0 d_{21}$
and from 2 and 3 we have
$d_{12}(d_{12}d_{21}) = d_{12}e = d_{12}$
therefore: $e_0 d_{21} = d_{12}$
multiplying on the left by e_0, we get from 4 and 3
$d_{21} = e_0 d_{12}$ ■

C4. $d_i(e_0 d_k) = (d_i e_0)d_k$
Proof: from 1, 4, and C2 we have:
$L = d_i(e_o d_k) = d_i((d_i d_i)d_k) = d_i(d_i(d_i d_k)) = (d_i d_i)(d_i d_k) = e_o(d_i d_k)$
$R = (d_i e_o)d_k = ((d_i(d_k d_k))d_k = ((d_i d_k)d_k)d_k = (d_i d_k)(d_k d_k) = (d_i d_k)e_o = e_o(d_i d_k)$ ■

C5. $d_i(ed_k) = (d_i e)d_k = e(d_i d_k)$
Proof: Trivial

C6. $(e_0d_i)(e_0d_i) = e_0$

Proof: from 4, 3, and C1, for $x = e_0d_i$, $y = e_0$, we have
$((e_0d_i)(e_0d_i))e_0 = (e_0d_i)((e_0d_i)e_0) = (e_0d_i)((d_ie_0)e_0) = (e_0d_i)(d_i(e_0e_0)) = (e_0d_i)d_i = e$
we thus have
$((e_0d_i)(e_0d_i))e_0 = e$
multiplying on the right by e_0, we get
$(((e_0d_i)(e_0d_i))e_0)e_0 = e_0$
therefore from 4 and 3 we have:
$e_0 = (((e_0d_i)(e_0d_i))e_0)e_0 = ((e_0d_i)(e_0d_i))(e_0e_0) = ((e_0d_i)(e_0d_i))e = (e_0d_i)(e_0d_i)$■

C7. $(e_0d_{12})(e_0d_{12}) = e_0$

Proof: the same as C6, for $x = e_0d_{12}$ and $y = e_0$■

C8. $(d_2d_1)(d_1d_2) = e$

Proof: $(d_2d_1)(d_1d_2) = (e_0d_{12})d_{12} =$ (C3)$= e$■

C9. $(d_2d_1)(d_2d_1) = e_0$

Proof: $(d_2d_1)(d_2d_1) = (e_0d_{12})(e_0d_{12}) =$ (C7) $= e_0$■

C1 - C9 allow us to construct Cayley table for discrimination in $G_{22}^{(0)}$.

TABLE FOR $G_{22}^{(0)}$

	d_1	d_2	d_{12}	e_0	e_0d_1	e_0d_2	e_0d_{12}	e
d_1	e_0	d_{12}	e_0d_2	e_0d_1	e	e_0d_{12}	d_2	d_1
d_2	e_0d_{12}	e_0	d_1	e_0d_2	d_{12}	e	e_0d_1	d_2
d_{12}	d_2	e_0d_1	e_0	e_0d_{12}	e_0d_2	d_1	e	d_{12}
e_0	e_0d_1	e_0d_2	e_0d_{12}	e	d_1	d_2	d_{12}	e_0
e_0d_1	e	e_0d_{12}	d_2	d_1	e_0	d_{12}	e_0d_2	e_0d_1
e_0d_2	d_{12}	e	e_0d_1	d_2	e_0d_{12}	e_0	d_1	e_0d_2
e_0d_{12}	e_0d_2	d_1	e	d_{12}	d_2	e_0d_1	e_0	e_0d_{12}
e	d_1	d_2	d_{12}	e_0	e_0d_1	e_0d_2	e_0d_{12}	e

2.2 Comment on the algebraic structure of $G_{22}^{(0)}$

(i) The first level, $G_{22}^{(0)}$, of NCH is the smallest dicyclic group (of order 8), isomorphic to the quaternion group, Q.

(ii) The subgroup of similarities of $G_{22}^{(0)}$, $S^{(0)}$, is a group spanned by the similarities, $\{e, e_0\}$, of $G_{22}^{(0)}$. $S^{(0)}$ is isomorphic to a cyclic group, C_2, of order 2. That is, $S^{(0)} \cong C_2$.

(iii) The subgroup of similarities, $S^{(0)}$, is the centre, $Z^{(0)}$, of $G_{22}^{(0)}$.

(iv) $S^{(0)}$ coincides with the derived group of $G_{22}^{(0)}$. That is
$S^{(0)} = [G_{22}^{(0)}, G_{22}^{(0)}] = \langle [x, y]: x, y \in G_{22}^{(0)} \rangle$

(v) For the group $G_{22}^{(0)}$, which is first level of NCH, the subgroup of similarities, the centre, and the derived group coincide:
$S^{(0)} = Z^{(0)} = [G_{22}^{(0)}, G_{22}^{(0)}]$

3. The Second Level of NCH

3.1 Generators of $G_{22}^{(1)}$

The second level, $G_{22}^{(1)}$, of NCH is generated by a set of automorphisms of $G_{22}^{(0)}$. The automorphisms are chosen in the following way:

(i) To each non-empty subset of the generators, $\{d_1, d_2\}$, of $G_{22}^{(0)}$ corresponds a subgroup of $G_{22}^{(0)}$ generated by the subset. That is, we consider the following subgroups of $G_{22}^{(0)}$: $\langle d_1 \rangle$, $\langle d_2 \rangle$, $\langle d_1, d_2 \rangle$. To each of these subgroups corresponds an automorphism of $G_{22}^{(0)}$ which leaves invariant each element of the subgroup.

(ii) We thus choose three automorphisms, $\{d_1^{(1)}, d_2^{(1)}, d_3^{(1)}\}$, each corresponding to a different subgroup of $G_{22}^{(0)}$ as indicated above. The automorphisms have to be chosen in such a way that none of them can be expressed as a sum of the others. Such automorphisms are called *independent.*

The above implies that the number of generators of the second level of NCH equals the number of subsets of the set of generators of the first level minus one (i.e., the empty subset of the generators). Thus, the number of generators of $G_{22}^{(1)}$ equals 2^2-1.

The following three independent automorphisms of $G_{22}^{(0)}$ are chosen[6] as the generative differences of $G_{22}^{(1)}$:

$$\mathbf{d_1^{(1)}}: (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (d_1, d_{12}, e_0d_2, e_0, e_0d_1, e_0d_{12}, d_2, e)$$
$$\mathbf{d_2^{(1)}}: (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (e_0d_{12}, d_2, d_1, e_0, d_{12}, e_0d_2, e_0d_1, e)$$
$$\mathbf{d_3^{(1)}}: (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e)$$

3.2 Discrimination in $G_{22}^{(1)}$

We define an operation on $\{d_1^{(1)}, d_2^{(1)}, d_3^{(1)}\}$ as the sum of automorphisms:

$$(d_i^{(1)}d_k^{(1)})(x) = d_i^{(1)}(x)d_k^{(1)}(x) \qquad \text{for all } x \in G_{22}^{(0)}$$

(where the operation on the right side is the discrimination operation in $G_{22}^{(0)}$)

The operation defined above is equivalent to that which C.W. Kilmister calls the 'induced discrimination' on the second level of NCH.

The automorphisms, $\{d_1^{(1)}, d_2^{(1)}, d_3^{(1)}\}$, with the induced operation of discrimination generate the second level, $G_{22}^{(1)}$, of NCH.

The generators of $G_{22}^{(1)}$ are automorphisms of $G_{22}^{(0)}$. However, because $G_{22}^{(0)}$ is not an abelian group, not all the elements of $G_{22}^{(1)}$ are automorphisms.

3.3 Algebraic structure of $G_{22}^{(1)}$

From the definition of discrimination on the generators of $G_{22}^{(1)}$, it follows:

$$(d_i^{(1)})^2: \quad (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (e_0, e_0, e_0, e, e_0, e_0, e_0, e)$$
$$(d_1^{(1)}d_2^{(1)})^2: \quad (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (e_0, e_0, e_0, e, e_0, e_0, e_0, e)$$
$$(d_1^{(1)}d_3^{(1)})^2: \quad (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (e, e_0, e_0, e, e, e_0, e_0, e)$$
$$(d_2^{(1)}d_3^{(1)})^2: \quad (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (e_0, e, e_0, e, e_0, e, e_0, e)$$
$$(d_1^{(1)}d_2^{(1)}d_3^{(1)})^2: \quad (d_1, d_2, d_{12}, e_0, e_0d_1, e_0d_2, e_0d_{12}, e) \rightarrow (e_0, e_0, e, e, e_0, e_0, e, e)$$

The above elements we call *similarities* of $G_{22}^{(1)}$, and denote them as follows:

$$\mathbf{e_0^{(1)}} = (d_i^{(1)})^2 \qquad i \in \{1, 2, 3\}$$
$$\mathbf{e_0^{(1)}} = (d_1^{(1)}d_2^{(1)})^2 = (d_{12}^{(1)})^2$$
$$\mathbf{e_1^{(1)}} = (d_1^{(1)}d_3^{(1)})^2 = (d_{13}^{(1)})^2$$
$$\mathbf{e_2^{(1)}} = (d_2^{(1)}d_3^{(1)})^2 = (d_{23}^{(1)})^2$$
$$\mathbf{e_3^{(1)}} = (d_1^{(1)}d_2^{(1)}d_3^{(1)})^2 = (d_{123}^{(1)})^2$$

In general, the similarities are the elements of $G_{22}^{(1)}$ which map every element of $G_{22}^{(0)}$ into the elements of $S^{(0)}$.

The similarities $\{e_0^{(1)}, e_1^{(1)}, e_2^{(1)}, e_3^{(1)}\}$ with the operation of (induced) discrimination span a group of order 8, which is called the *group of similarities*, $S^{(1)}$, of the second level of NCH.

Since the order of each element of $S^{(1)}$ is 2, $S^{(1)}$ is an abelian group. It is isomorphic to a direct product of cyclic groups (of order 2): $C_2 \times C_2 \times C_2$.

The identity of $S^{(1)}$ is denoted by $e^{(1)} = e_1^{(1)} e_2^{(1)} e_3^{(1)}$. The identity maps every element of $G_{22}^{(0)}$ into $\{e\}$.

From the fact that $G_{22}^{(0)}$ is a non-commutative group it follows that $G_{22}^{(1)}$ is a non-commutative group. The subgroup of similarities, $S^{(1)}$, is the centre of $G_{22}^{(1)}$, and the identity of $S^{(1)}$ is the identity of $G_{22}^{(1)}$.

The following relationships which hold for the elements of $G_{22}^{(1)}$ let us construct the Cayley table for $G_{22}^{(1)}$:

$e^{(1)} = d_{ik}^{(1)} d_{ki}^{(1)}$ for $i \neq k$ and $i, k \in \{1, 2, 3\}$
$e_0^{(1)} = (d_i^{(1)})^2 = (d_{12}^{(1)})^2$ $i \in \{1, 2, 3\}$
$e_1^{(1)} = (d_{13}^{(1)})^2$
$e_2^{(1)} = (d_{23}^{(1)})^2$
$e_3^{(1)} = (d_{123}^{(1)})^2 = e_1^{(1)} e_2^{(1)} = e_{12}^{(1)}$

$d_{21}^{(1)} = e_0^{(1)} d_{12}^{(1)}$
$d_{31}^{(1)} = e_1^{(1)} d_{13}^{(1)}$
$d_{32}^{(1)} = e_2^{(1)} d_{23}^{(1)}$
$d_{321}^{(1)} = e_0^{(1)} e_3^{(1)} d_{123}^{(1)}$
$d_{321}^{(1)} d_{123}^{(1)} = e_0^{(1)}$

Table for $G_{22}^{(1)}$. Part O

	$d_1^{(1)}$	$d_2^{(1)}$	$d_{12}^{(1)}$	$d_3^{(1)}$	$d_{13}^{(1)}$	$d_{23}^{(1)}$	$d_{123}^{(1)}$	$e^{(1)}$
$d_1^{(1)}$	$e_0^{(1)}$	$d_{12}^{(1)}$	$e_0^{(1)}d_2^{(1)}$	$d_{13}^{(1)}$	$e_0^{(1)}d_3^{(1)}$	$d_{123}^{(1)}$	$e_0^{(1)}d_{23}^{(1)}$	$d_1^{(1)}$
$d_2^{(1)}$	$e_0^{(1)}d_{12}^{(1)}$	$e_0^{(1)}$	$d_1^{(1)}$	$d_{23}^{(1)}$	$e_0^{(1)}d_{123}^{(1)}$	$e_0^{(1)}d_3^{(1)}$	$d_{13}^{(1)}$	$d_2^{(1)}$
$d_{12}^{(1)}$	$d_2^{(1)}$	$e_0^{(1)}d_1^{(1)}$	$e_0^{(1)}$	$d_{123}^{(1)}$	$d_{23}^{(1)}$	$e_0^{(1)}d_{13}^{(1)}$	$e_0^{(1)}d_3^{(1)}$	$d_{12}^{(1)}$
$d_3^{(1)}$	$e_1^{(1)}d_{13}^{(1)}$	$e_2^{(1)}d_{23}^{(1)}$	$e_{12}^{(1)}d_{123}^{(1)}$	$e_0^{(1)}$	$e_{01}^{(1)}d_1^{(1)}$	$e_{02}^{(1)}d_2^{(1)}$	$e_{012}^{(1)}d_{12}^{(1)}$	$d_3^{(1)}$
$d_{13}^{(1)}$	$e_{01}^{(1)}d_3^{(1)}$	$e_2^{(1)}d_{123}^{(1)}$	$e_{012}^{(1)}d_{23}^{(1)}$	$e_0^{(1)}d_1^{(1)}$	$e_1^{(1)}$	$e_{02}^{(1)}d_{12}^{(1)}$	$e_{12}^{(1)}d_2^{(1)}$	$d_{13}^{(1)}$
$d_{23}^{(1)}$	$e_{01}^{(1)}d_{123}^{(1)}$	$e_{02}^{(1)}d_3^{(1)}$	$e_{12}^{(1)}d_{13}^{(1)}$	$e_0^{(1)}d_2^{(1)}$	$e_1^{(1)}d_{12}^{(1)}$	$e_2^{(1)}$	$e_{012}^{(1)}d_1^{(1)}$	$d_{23}^{(1)}$
$d_{123}^{(1)}$	$e_1^{(1)}d_{23}^{(1)}$	$e_{02}^{(1)}d_{13}^{(1)}$	$e_{012}^{(1)}d_3^{(1)}$	$e_0^{(1)}d_{12}^{(1)}$	$e_{01}^{(1)}d_2^{(1)}$	$e_2^{(1)}d_1^{(1)}$	$e_{12}^{(1)}$	$d_{123}^{(1)}$
$e^{(1)}$	$d_1^{(1)}$	$d_2^{(1)}$	$d_{12}^{(1)}$	$d_3^{(1)}$	$d_{13}^{(1)}$	$d_{23}^{(1)}$	$d_{123}^{(1)}$	$e^{(1)}$

Table for $G_{22}^{(1)}$

O	$e_1^{(1)}$O	$e_2^{(1)}$O	$e_{12}^{(1)}$O	e_0O	$e_{01}^{(1)}$O	$e_{02}^{(1)}$O	$e_{012}^{(1)}$O
$e_1^{(1)}$O	O	$e_{12}^{(1)}$O	$e_2^{(1)}$O	$e_{01}^{(1)}$O	e_0O	$e_{012}^{(1)}$O	$e_{02}^{(1)}$O
$e_2^{(1)}$O	$e_{12}^{(1)}$O	O	$e_1^{(1)}$O	$e_{02}^{(1)}$O	$e_{012}^{(1)}$O	e_0O	$e_{01}^{(1)}$O
$e_{12}^{(1)}$O	$e_2^{(1)}$O	$e_1^{(1)}$O	O	$e_{012}^{(1)}$O	$e_{02}^{(1)}$O	$e_{01}^{(1)}$O	e_0O
e_0O	$e_{01}^{(1)}$O	$e_{02}^{(1)}$O	$e_{012}^{(1)}$O	O	$e_1^{(1)}$O	$e_2^{(1)}$O	$e_{12}^{(1)}$O
$e_{01}^{(1)}$O	e_0O	$e_{012}^{(1)}$O	$e_{02}^{(1)}$O	$e_1^{(1)}$O	O	$e_{12}^{(1)}$O	$e_2^{(1)}$O
$e_{02}^{(1)}$O	$e_{012}^{(1)}$O	e_0O	$e_{01}^{(1)}$O	$e_2^{(1)}$O	$e_{12}^{(1)}$O	O	$e_1^{(1)}$O
$e_{012}^{(1)}$O	$e_{02}^{(1)}$O	$e_{01}^{(1)}$O	e_0O	$e_{12}^{(1)}$O	$e_2^{(1)}$O	$e_1^{(1)}$O	O

3.4 Comment on the algebraic structure of $G_{22}^{(1)}$

(i) The second level, $G_{22}^{(1)}$, of NCH is a non-commutative group of order 64. It is a finite 2-group, and its nilpotency class is 2.

(ii) The subgroup of similarities, $S^{(1)}$, of $G_{22}^{(1)}$ is the centre, $Z^{(1)}$, of $G_{22}^{(1)}$. It is a direct product of cyclic groups of order 2. That is

$S^{(1)} \cong C_2 \times C_2 \times C_2$ (for example $S^{(1)} \cong \langle e_0 \rangle \times \langle e_1 \rangle \times \langle e_2 \rangle$)

(iii) $S^{(1)}$ coincides with the derived group of $G_{22}^{(1)}$

$S^{(1)} = [G_{22}^{(1)}, G_{22}^{(1)}]$

(iv) For the group $G_{22}^{(1)}$, which is the second level of NCH, the subgroup of similarities, the centre, and the derived group coincide:

$S^{(1)} = Z^{(1)} = [G_{22}^{(1)}, G_{22}^{(1)}]$

4. On the Relationship of NCH and CCH

(i) The tables (1) and (2) display the commutative (CCH) and the noncommutative (NCH) combinatorial hierarchies. G-differences refer to generative differences, and G-similarities refer to generative similarities of appropriate levels.

TABLE 1: CCH

LEVELS	NUMBER OF G-DIFFERENCES	CCH	$\lvert G_{12}^{(i)}\rvert = 2n_i$
0	$n_0 = 2$	$G_{12}^{(0)}$	2^2
1	$n_1 = 2^2-1 = 3$	$G_{12}^{(1)}$	2^3
2	$n_2 = 2^3-1 = 7$	$G_{12}^{(2)}$	2^7
3	$n_3 = 2^7-1 = 127$	$G_{12}^{(3)}$	2^{127}
4	$n_4 = 2^{127}-1$	$G_{12}^{(4)}$	$2^{2^{127}-1}$
i	$n_{i+1} = 2^{n_i}-1$		2^{n_i}

TABLE 2: NCH

LEV.	NCH	NUMBER OF G-DIFFERENCES	NUMBER OF G-SIMILARITIES	$S^{(i)}$	$\lvert S^{(i)}\rvert$	$\lvert G_{22}^{(i)}\rvert = 2^{n_i}\lvert S^{(i)}\rvert$
0	$G_{22}^{(0)}$	$n_0 = 2$	1	$S^{(0)}$	2	2^{2+1}
1	$G_{22}^{(1)}$	$n_1 = 3$	3	$S^{(1)}$	2^3	2^{3+3}
2	$G_{22}^{(2)}$	$n_2 = 7$	3×7	$S^{(2)}$	2^{21}	2^{7+21}
3	$G_{22}^{(3)}$	$n_3 = 127$	127× 63	$S^{(3)}$		
4	$G_{22}^{(4)}$	$n_4 = 2^{127}-1$		$S^{(4)}$		
i	$G_{22}^{(i)}$	$n_{i+1} = 2^{n_i} - 1$	$k_i = n_i(n_i-1)/2$		$2^{k_i} = 2^{n_i(n_i-1)/2}$	$2^{n_i+k_i} = 2^{n_i(n_i+1)/2}$

TABLE 3: RELATIONSHIP OF NCH AND CCH

LEVELS	ISOMORPHISM
0	$G_{22}^{(0)}/S^{(0)} \cong G_{12}^{(0)}$
1	$G_{22}^{(1)}/S^{(1)} \cong G_{12}^{(1)}$
2	$G_{22}^{(2)}/S^{(2)} \cong G_{12}^{(2)}$
3	see (iii)
4	see (iii)

(ii) The table (3) shows a relationship between the levels of the commutative and non-commutative combinatorial hierarchies. The relationship is the following: $G_{22}^{(i)}/ S^{(i)} \cong G_{12}^{(i)}$, and it holds for $i \in \{0, 1, 2\}$.[7]

That is, for these levels the CCH is isomorphic to a quotient group of $G_{22}^{(i)}$ by the subgroup of similarities, $S^{(i)}$, of the corresponding level of NCH.

Thus, each of the first three levels of the CCH is isomorphic to a homomorphic image of a corresponding level of NCH, when we consider the following homomorphisms: $H_i : G_{22}^{(i)} \rightarrow G_{22}^{(i)}/ S^{(i)} \cong G_{12}^{(i)}$

For these levels of the NCH, the subgroups of similarities, the centres, (and the derived groups) coincide. Thus, the levels of CCH are isomorphic to the groups of *inner automorphisms* of the corresponding levels of NCH.

$I(G_{22}^{(i)}) \cong G_{12}^{(i)}$

where $I(G_{22}^{(i)})$ is the group of inner automorphisms of $G_{22}^{(i)}$ under *product* of the automorphisms.

(iii) For i = 3, the order of the subgroup of similarities, $S^{(3)}$, is greater than the number of mappings: $G_{22}^{(2)} \rightarrow S^{(2)}$ which *limits* the order of $S^{(3)}$.

That is, on one hand, the order of $S^{(i)}$ is $2^{k_i} = 2^{n_i(n_i-1)/2}$, (and this holds for i = 0, 1, 2), on the other hand, from the definition of the successive levels of NCH it follows that the number of similarities, $|S^{(i)}|$, is not greater than $|S^{(i-1)}|$ to the power of $(2^{n_{i-1}} - 1)$.[8]

In other words, for i = 3, we should have

$|S^{(3)}| = 2^{127\times 63}$ and $|S^{(3)}| \leq 2^{3\times 7\times 127}$

which is a contradiction.

The above implies that for i = 3 not all similarities are different. What bearing this has for the ending of NCH will be the subject of the next paper.

NOTES:

[1] D. Bohm, *Wholeness and the mmplicate Order*, Routledge & Kegan Paul, 1980

[2] T. Bastin, C.W. Kilmister, *Combinatorial Physics*, World Scientific, 1995

[3] C.W. Kilmister, *Discrimination with Aspect*, manuscript 1977

[4] This is a rather partial response to Clive's reservations against associativity. However, it is still interesting that assuming the left and right alternative laws of associativity only one may dispense with Conway's method.

[5] The same notation is used throughout the paper (i.e., d_{ij} denotes $d_i d_j$, d_{ijk} denotes $d_i d_j d_k$, etc.).

[6] As C.W. Kilmister demonstrated, [3], these three automorphisms are uniquely determined.

[7] Although the relationship was demonstrated only for i = 0, 1, it is possible to show, using the same reasoning as for passing from the level 0 to the level 1, that it still holds for i = 2.

[8] This calculation is based on the fact that the similarities of $G_{22}^{(i)}$ are mappings: $G_{22}^{(i-1)} \rightarrow S^{(i-1)}$, and the number of these mappings equals the number of mappings of all non-empty subsets of generators of $G_{22}^{(i-1)}$ into $S^{(i-1)}$.

A COMBINATORIAL BIT BANG LEADING TO QUATERNIONS[1]

by
Michael Manthey
Department of Mathematics and Computer Science
Aalborg University
Frederik Bajersvej 7E
9220 Aalborg DENMARK
manthey@iesd.auc.dk

Abstract. This paper describes in detail how (discrete) quaternions - ie. the abstract structure of 3-D space - emerge from, first, the Void, and thence from primitive combinatorial structures, using only the exclusion and co-occurrence of otherwise unspecified events. We show how this computational view supplements, and provides an interpretation for, the mathematical structures. The build-up is emergently hierarchical, compatible with both quantum mechanics and relativity, and can be extended upwards to the macroscopic. The mathematics is that of Clifford algebras emplaced in the homology-cohomology structure pioneered by Kron. Interestingly, the ideas presented here were originally developed by the author to resolve fundamental limitations of existing artificial intelligence paradigms.

1 Introduction

We find ourselves in a universe of myriad, mystifying, and very nearly incomprehensible, complexity. At the same time, contemporary Big Bang cosmogenesis tells us that this complexity has apparently *emerged* from 'nothing' - from *Void* - via a poorly understood process. In this paper, I will attempt to describe a discrete, combinatorial, and computational framework for this process. I gladly acknowledge the inspiration of The Combinatorial Hierarchy of [Bastin&Kilmister, Parker-Rhodes], although the material presented here differs herefrom in many ways.

How can one get something from nothing? This question is currently being framed in terms of the concept of *emergence* - that novel properties can emerge from simpler constituents, while simultaneously these properties cannot be reduced to isolated actions of said constituents. From our point of view, the concept of emergence cannot be separated from that of *hierarchy*, in that emergent properties by definition inhabit a 'higher' - that is, more complex - level of organization than their constituents. The fact that the concept of emergence is controversial is, in our view, a result of the residual, but all-pervasive, influence of Newton's physics, which is entirely reductionistic.

Nevertheless, a great advantage of the Newtonian view is that it provides an intuitive *mechanism* for how material entities influence each other: momentum exchange, as in

[1] To appear in the Proceedings of the 1997 Helsinki Conference on Emergence, Complexity, Hierarchy, Organization (ECHO III).

the universal image of billiard balls colliding and rebounding. The felicity of mechanism is that it provides a blow-by-blow description of what is going on, and in so doing fertilizes our imagination while simultaneously guiding our modelling choices. The quantum-mechanical revolution put an end to this, evaporating 'materia' into a cloud of probability amplitudes, uncertainty, and non-determinism. Lacking the compass of a trustworthy mechanism, we have been collectively doomed to purblindly wander the jungles of mathematics, as Einstein well appreciated. For mathematics, in spite of appearances, does not really describe *how* things happen, but rather only their essential *what.*

Although we do not argue the case particularly here (see [Manthey94] for some initial conjectures), our model implicitly supplies an *informational* mechanism for quantum mechanics, and at that, one that is compatible with relativity theory. The mechanism we present is computational, in the sense that it is discrete, combinatorial in character, and described in terms of discrete computational operations. However, these operations are not the usual arithmetic "number crunching" most people associate with computing. Nor is the computation in question describable in terms of a Turing machine, which is - as [Penrose] essentially argued - equivalent to Newtonian mechanics. Rather, the concept of an evolving and expanding universe demands a *distributed multi-process* view. Thus the critical mechanisms are those that express the *synchronization* between the events constituting the various processes. The model that is built up on this basis, and explained in the following, we have dubbed the *phase web paradigm*; a corresponding program, called *Topsy*, has been implemented and is available to interested persons [www].

A major contribution of this paper is therefore that it shows how to unite computational, *informational* mechanisms - with the aforementioned advantages hereof - with 'classical' vector algebra, with surprising and intriguing results. In addition, the hierarchical aspect of the basic mechanisms shows how it is possible, at least in principle, to tell a detailed and rigorous story about the ascent from the microscopic to the macroscopic world.

The outline of the paper is as follows: the next section sketches our conceptual, and decidedly computational, framework, its mapping to Clifford algebras and a novel hierarchical structure that naturally captures emergent phenomena. The following section connects this with our implementation, revealing an important ambiguity in the mathematical description versus concrete *informational* mechanism, which ambiguity is then resolved. We then present a combinatorial "bit bang" based on the mechanisms introduced, and show how quaternions appear.

2 Mechanism, Clifford algebra, and Hierarchy

Initially, it is crucial to establish the validity of the concept of emergence in a mechanistic context.[2] We will see that the presence of multiple processes is critical for this purpose.

***The coin demonstration** - Act I. A man stands in front of you with both hands behind his back, whilst you have one hand extended in front of you, palm up. You see the man move one hand from behind his back and place a coin on your palm. He then removes the coin with his hand and moves it back behind his back. After a brief pause, he again moves his hand from behind his back, places what appears to be an identical coin in your palm, and removes it again in the same way. He then asks you, "How many coins do I have?".*

It is important at the outset to understand that the coins are *formally* identical: indistinguishable in every respect. If you are unhappy with this, replace them with electrons or geometric points. Also, there are no 'tricks' in the prose formulation. What is at issue is the fact of indistinguishability, and we are simply trying to pose a very simple situation where it is indistinguishability, and nothing else, that is in focus.

The indistinguishability of the coins now agreed, the most inclusive answer to the question is "One or more than one", an answer that exhausts the universe of possibilities given what you have seen, namely *at least* one coin. There being exactly two possibilities, the outcome can be encoded in one bit of information. Put slightly differently, when you learn the answer to the question, you will per force have received one bit of information.

***The coin demonstration** - Act II. The man now extends his hand and you see that there are two coins in it. [The coins are of course identical.]*

You now know that there are two coins, that is, *you have received one bit of information.* We have now arrived at the final act in our little drama.

***The coin demonstration** - Act III. The man now asks, "Where did that bit of information come from??"*

Indeed, where *did* it come from?! Since the coins are indistinguishable, seeing them one at a time will never yield an answer to the question. Rather, *the bit originates in the simultaneous presence of the two coins.* We call such a confluence a *co-occurrence.*[3]

Penrose [Penrose] has argued that computational systems, not least parallel ditto, *in principle* cannot model quantum mechanics. However, his argument is based on Turing's model, which in turn cannot capture co-occurrence.

Notice by the way how the matrix-based formulations of QM neatly get around the inherent sequentiality of $y = f(x)$-style (ie. algorithmic) thinking, namely by the literal

[2] Since we are dealing with informational mechanism, we prefer the term *neo*-mechanistic.

[3] At this juncture, we hasten to mention that we are dealing here with *local* simultaneity, so there is no collision with relativity theory. Indeed, Feynman [Feynman65 p.63] argues from the basic principle of relativity of motion, and thence Einstein locality, that if *anything* is conserved, it must be conserved *locally.*

co-occurrence of values in its vectors' and matrices' very layouts; and thereafter by how these values are composed *simultaneously* (conceptually speaking) by matrix operations. Instead of the matrix route, we have taken the conceptually compatible one of Clifford algebras, which are much more compact, elegant, and general, cf. [Hestenes].

We see from the Coin demonstration that there is information, *computational information*, available in the universe ***which in principle cannot be obtained sequentially.*** One can say that the information received from observing a co-occurrence is indicative of the fact that two states do not mutually exclude each other.

Co-occurrence and mutual-exclusion are in fact conceptual *opposites*, in that (say) two events cannot simultaneously both co-occur and mutually exclude. The following shows how this insight can be promoted to a concept of 'action'.

The block demonstration. *Imagine two 'places', p and q, each of which can contain a single 'block'. Each of the places is equipped with a sensor, s_p respectively s_q, which can indicate the presence or absence of a block.*

The sensors are the *only* source of information about the state of their respective places and are assumed *a priori* to be independent of each other, though they may well be correlated. The two states of a given sensor s are mutually exclusive, so a place is always either 'full', denoted (arbitrarily) by s, or 'empty', denoted by $\tilde{s}$; clearly, $\tilde{\tilde{s}} = s$.[4]

Suppose there is a block on p and none on q. This will allow us to observe the co-occurrence $s_p + \tilde{s}_q$. From this we learn that having a block on p does not exclude not having a block on q. Suppose at some other instant (either before or after the preceding) we observe the opposite, namely $\tilde{s}_p + s_q$. We now learn that not having a block on p does not exclude having a block on q. What can we conclude?

First, it is important to realize that although the story is built around the co-occurrences $s_p + \tilde{s}_q$ and $\tilde{s}_p + s_q$, everything we say below applies equally to the 'dual' pair of co-occurrences $s_p + s_q$ and $\tilde{s}_p + \tilde{s}_q$. After all, the designation of one of a sensor's two values as '$\sim$' is entirely arbitrary. It is also important to realize that the places and blocks are story props: all we really have is two two-valued sensors reflecting otherwise unknown activities in the surrounding environment. Such sensors constitute the *boundary* between an entity and its environment in the phase web paradigm.

Returning to the question posed, we know that s_p excludes $\tilde{s}_p$ and similarly s_q excludes $\tilde{s}_q$. Furthermore, we have observed the co-occurrence of s_p and $\tilde{s}_q$ and vice versa. Since the respective parts of one co-occurrence exclude their counterparts in the other co-occurrence (cf. first sentence), we can conclude that the co-occurrences *as wholes* exclude each other.

[4] We are working in $\mathbb{Z}_3 = \{0, 1, -1 = \tilde{1}\}$ rather than the traditional $\mathbb{Z}_2 = \{0, 1\}$. We use the visual convention that a sensor written without a tilde is taken to be bound to the value 1, and vice versa; clearly, $\tilde{0} = 0$.

Take this now a step further. The transition $s_p \rightarrow \tilde{s}_p$ is indicative of some *action* in the environment, as is the reverse, $\tilde{s}_p \rightarrow s_p$. The same applies to s_q. Perceive the transitions $s_p \leftrightarrow \tilde{s}_p$ and $s_q \leftrightarrow \tilde{s}_q$ as two sequential computations, each of whose states consists of a single value-alternating bit. By the independence of sensors, these two computations are completely independent of each other. At the same time, the logic of the preceding paragraph allows us to infer the existence of a third computation, a *compound* action, with the state transition $s_p + \tilde{s}_q \leftrightarrow \tilde{s}_p + s_q$, denoted $s_p s_q$. In effect, by combining in this way two single-bit computations to yield one two-bit computation, we have lifted our conception of the actions performable by the environment to a new, higher, level of abstraction. This inference we call *co-exclusion*, and can be applied to co-occurrence pairs of any arity > 1 where at least two corresponding components have changed.

Notice that the same reasoning applies to the action $s_p + s_q \leftrightarrow \tilde{s}_p + \tilde{s}_q$, also denoted $s_p s_q$. The two actions are, not surprisingly, *dual* to each other, so co-exclusion on two sensors can generate two distinct actions. Like co-occurrence, an action defined by co-exclusion also possesses an emergent property, generally comparable to spin $\frac{1}{2}$ [Manthey94].

Co-exclusion provides a very general mechanism for self-organization: simply observe co-excluding co-occurrences, since these then will represent an abstraction of the environment. However, the mechanism for actually discovering co-exclusions is as yet unspecified. Speaking mechanistically, how *exactly* does one (eg. the universe) discover the existence of a co-exclusive relationship between two co-occurrences?

Define a co-occurrence in terms of an "event buffer" with time-window-size Δt, where true simultaneity requires that $\Delta t = 0$, and larger values recognize the factual granularity with one can resolve events and/or the time-scale at which an environment varies. Suppose further that event identifiers are put into the event buffer as they occur, ie. the new state engendered (and labelled) by the action associated with the event is inserted into the buffer. Finally, suppose that events in the buffer are successively discarded as their residence exceeds Δt or the same event-state changes again. Clearly, this arrangement guarantees that the state changes contained in the buffer all took place within Δt, and thus occurred 'simultaneously' (modulo Δt). The reader is at this point encouraged to ponder the fact that this mechanism in fact solves the problem of discovering co-exclusions, and at that, in linear time and space and without pre-specification! [Reader pause, for a lovely *aha!* experience.]

To see why this claim is true, consider the fact that a sensor's states are mutually exclusive, that is, if a sensor is currently in state s then before it changed it was in the state $\tilde{s}$. Furthermore, in $\mathbb{Z}_3$ at least, the opposite is also true: $\tilde{\tilde{s}} = s$. Hence, since the buffer contains the co-occurrence (say) $s_1 + s_2$, and *they both just changed*, then before they entered the buffer, $\tilde{s}_1 + \tilde{s}_2$ obtained. But these two co-occurrences are exactly those necessary to define the co-exclusion $s_1 + s_2 \longleftrightarrow \tilde{s}_1 + \tilde{s}_2$! The computation time and space are fundamentally linear because they are proportional to the buffer size. If we specify that *all* events are to pass through our event buffer, then the only pre-specification is the arity of the co-exclusion. Even this pre-specification can be avoided if all possible co-exclusions over the current buffer contents are instantiated as each event is entered into the buffer.

2.1 Co-occurrence and Co-exclusion via Clifford Algebras

This section presents, very informally, the mathematical foundation of the phase web paradigm. The point of departure is to view sensor states as vectors instead of scalars, as is conventionally done.

Let sensor state $s = 1$ indicate that sensor s is currently being stimulated, ie. a synchronization token (an informational marker for a state's existence) for that state is present, and $s = \tilde{1}$ that s is currently *not* being stimulated, and hence a token for state $\tilde{s}$ is present. Thus the two states of s are represented by their respective synchronization tokens, whose respective presences by definition exclude each other.

That a set of sensors *qua* vectors are orthogonal derives from the fact that, in principle, a given sensor says nothing about the state of any other sensor. A state of a multi-sensor system is then naturally expressed as the sum of the individual sensor vectors, and the state $(s_a, \tilde{s}_b) = (1, \tilde{1})$ is written as the vector sum $s_a + \tilde{s}_b$. Since such states represent co-occurrences, it follows that co-occurrences are vector sums, usually denoting partial (local) states. Note how the commutativity of '+' reflects the lack of ordering of the components of a co-occurrence; and as well that the co-occurrence $1 + \tilde{1} = 0$ indicates that the interpretation of 'zero' is that the components of the sum *exclude* each other. Because $\mathbb{Z}_2$ does not distinguish state-value and exclusion, we take our algebra to be over $\mathbb{Z}_3 = \{0, 1, 2\} = \{0, 1, \tilde{1}\}$.

The next step is to represent *actions*. [Manthey94] presents a detailed analysis of the group properties of co-occurrences and actions, concluding that the appropriate algebraic formalism is a (discrete) Clifford algebra [Hestenes], and that the state transformation effected by an action is naturally expressed using this algebra's vector product. A prime characteristic of this product is that it is anti-commutative, that is, for $(s_1)^2 = (s_2)^2 = 1$, $s_1 s_2 = -s_2 s_1$.[5] The magnitude of any such product is the area of the parallelogram its two components span, and the *orientation* of the product is perpendicular to the plane of the parallelogram and determined by the "right hand rule". Applying the Clifford product to a state, one finds - using the square-rule and the anti-commutativity of the product given above - that

$$(s_1 + s_2)s_1 s_2 = s_1 s_1 s_2 + s_2 s_1 s_2 = s_2 + \tilde{s}_1 s_2 s_2 = \tilde{s}_1 + s_2 \qquad (1)$$

[5]The Clifford product ab can be defined as $ab = a \cdot b + a \wedge b$, ie. the sum of the inner ($\cdot$) and outer ($\wedge$) products, where $a \wedge b = -b \wedge a$ is the oriented area spanned by vectors a, b. The basis vectors s_i of a Clifford algebra may have $(s_i)^2 = \pm 1$, and while here we choose $+1$, reasons are appearing for choosing -1. As long as they all have the same square, it doesn't matter for what is said here. Note that $(s_1 s_2)^2 = -1$, so $s_1 s_2 \cong \sqrt{-1}$.

that is, that the result of the *action* s_1s_2 is to rotate the original state by 90°, for which reason things like s_1s_2 are called *spinors*. Thus *state change* in the phase web is modelled by rotation (and reflection) of the state space, and the effect of an 'entire' action can be expressed by the inner automorphism $s_1s_2(s_1 + s_2)s_2s_1 = \bar{s}_1 + \bar{s}_2$, which corresponds to a rotation through 180°.

One of the felicities of Clifford algebras is that one needn't designate one of the axes as 'imaginary' and the others as 'real'. Rather, the i-business is implicit and the algebra's anti-commutative product neatly bookkeeps the desired orthogonality and inversion relationships, no matter how many dimensions [ie. sensors (roughly)] are present.

The above 2-spinors are just one example of the vector products available in a Clifford algebra - any product of the basis vectors s_i is well-defined, and just as s_1s_2 defines an area, $s_1s_2s_3$ defines a volume, etc. Being by nature mutually orthogonal, the terms of a Clifford algebra

$$s_i + s_is_j + s_is_js_k + \ldots + s_is_j \ldots s_n \tag{2}$$

themselves also define a vector space, which is the space in which we will be working (actually, hierarchies of such spaces). [The term (eg.) s_is_j above, for $n = 3$, denotes $s_1s_2 + s_2s_3 + s_3s_1$, that is, all possible non-redundant combinations.] It is perhaps worth stressing that this vector space is the space of the *distinctions* expressed by sensors, and as such has no direct relationship with ordinary 3+1 dimensional space.

A Clifford product like s_1s_2 reflects both (1) the emergent aspect of a phase web action (via its perpendicularity to its components) and (2) its ability to act as a meta-sensor (since its orientation is ± 1). Regarding (1), the emergence is rooted in the information gleaned from the co-occurrences underlying the co-exclusion inference that yields s_1s_2, cf. the Coin demonstration. Regarding (2), the co-exclusion inference is an *abstraction* that produces a single action with two bits of state from two lower level actions each possessing a single bit of state. Since this abstraction has the same external behavior as its constituent sensors, namely ± 1, we can legitimately view it too as a sensor, a *meta*-sensor. By co-excluding meta-sensors, we can build a new set of abstractions - meta-meta-sensors - etc., and thus construct a hierarchy of interwoven co-occurrences and exclusions that directly reflects the *observed* activity of the surrounding environment. This hierarchy is the topic of the following.

2.2 From Clifford Algebra to Hierarchy

In analogy to s_1, s_2 co-excluding to yield s_1s_2, one might expect that the co-exclusion of two meta-sensors, say s_is_j and s_ps_q, would be modelled by simply multiplying them, to get the 4-action $s_is_js_ps_q$. This turns out however to be inadequate, since although by the same logic the co-exclusion of (say) s_i and s_is_j in a phase web expresses explicitly a useful relationship (eg. part-whole), the algebra's rules reduce it from $s_is_is_j$ to s_j, which is simply redundant.

Instead, we take as a clue the fact that *change* in a phase web occurs via trickling down through the layers of hierarchy, and draw an analogy with differentiation. In the present decidedly geometric and discrete context, differentiation corresponds to the *boundary operator* ∂. Define $\partial s = 1$ and let

$$\partial(s_1 s_2 \ldots s_m) = s_2 s_3 \ldots s_m - s_1 s_3 \ldots s_m + s_1 s_2 s_4 \ldots s_m - \ldots (-1)^{m+1} s_1 s_2 \ldots s_{m-1}$$

that is, drop one component at a time, in order, and alternate the sign.[6] Using the algebra's rules as before, one can show that $\partial(s_1 s_2 \ldots s_m) = (s_1 + s_2 + \ldots + s_m) s_1 s_2 \ldots s_m$ which is exactly the form of equation (1) for what an action does!

Take now equation (2) expressing the vector space of distinctions, segregate terms with the same arity, and arrange them as a decreasing series:

$$s_i \xleftarrow{\partial} s_i s_j \xleftarrow{\partial} s_i s_j s_k \xleftarrow{\partial} \ldots \xleftarrow{\partial} s_i s_j \ldots s_{n-1} \xleftarrow{\partial} s_i s_j \ldots s_n \qquad (3)$$

Here as before, $s_i s_j$ is to be understood as expressing all the possible 2-ary forms (etc.), and hence the co-occurrence of pieces of similar structure. Each of the individuals is a *simplicial complex*, and the whole sequence is called a *chain complex*, expressing a sequence of structures of graded geometrical complexity in which the transition from a higher to a lower grade is defined by ∂. Furthermore, the entities at adjacent levels are related via their group properties - their *homology*, which we here assume is trivial.

The basic mechanism for expressing change or action in our hierarchical context is that of *goal-driven* computation. A *goal* is a local state whose presence causes an action to attempt to change its orientation, and a goal will typically be decomposed recursively into subgoals on that action's constituents as it trickles down through the ∂-hierarchy. [Goals differ from the 'imperatives' traditionally used in computing - eg. add x,2 or sine(x) - by not guaranteeing that the indicated computation will be achieved, but rather only a 'best effort', and success is contingent on the state of the environment and the rest of the phase web. There is no teleological baggage per se in this concept - 'potential' is a closer idea.]

It turns out that there is a second structure - a *cohomology* - that is isomorphic to the homology, but with the difference that arity *increases* via the δ (or *co-boundary*) operator,[7] precisely opposite to ∂, cf. eqn. (3):

$$s_i \xrightarrow{\delta} s_i s_j \xrightarrow{\delta} s_i s_j s_k \xrightarrow{\delta} \ldots \xrightarrow{\delta} s_i s_j \ldots s_{n-1} \xrightarrow{\delta} s_i s_j \ldots s_n \qquad (4)$$

Building such increasing complexity is exactly what co-exclusion does. [We note that a Clifford algebra satisfies the formal requirements for the existence of the associated homology and cohomology.]

[6] If one takes two components at a time, as we will do on occasion later on, then the sign-alternation disappears.

[7] More precisely, $(\sigma_p, \delta d^{p-1}) = (\sigma_p \partial, d^{p-1})$, where σ_p is a simplicial complex with arity p, and d^p the corresponding co-complex.

It is easily proven that $\partial\partial = 0$, and by isomorphism, so also $\delta\delta = 0$. For example, $\partial\partial(s_1 s_2) = \partial(\tilde{s}_1 + s_2) = \tilde{1} + 1 = 0$, and similarly, $\partial\partial(s_2 s_1) = \partial(s_1 + \tilde{s}_2) = 1 + \tilde{1} = 0$. Combining these now as the exclusion $\partial\partial(s_1 s_2 + s_2 s_1)$, we get $(1 + \tilde{1}) + (\tilde{1} + 1) = (1 + 1) + (\tilde{1} + \tilde{1}) = 0$, which are the two forms of the input to the determination of a co-exclusion relationship. Recalling the event-buffer mechanism for discovering co-exclusions, we see, especially if $\Delta t = 0$, that this mechanism is a realization of the isomorphic $\delta\delta = 0$!

Viewing δ's abstraction operation informationally, we see that two bits (s_1, s_2) are being encoded in a single bit (the orientation of $s_1 s_2$), that is, information is being 'abstracted away'. The missing bit indicates the *phase* of the action, ie. whether the state rotation/transformation is $s_1 + s_2 \leftrightarrow \tilde{s}_1 + \tilde{s}_2$ or $s_1 + \tilde{s}_2 \leftrightarrow \tilde{s}_1 + s_2$. What will actually occur is however well-defined by the other connections s_1, s_2 partake in, ie. the boundary conditions of the action. Note however that 'well-defined' does not necessarily imply 'deterministic'. Isomorphically, the corresponding ∂ operation destroys the emergent information in the current state and replaces it by non-deterministic outcome.

Refer now to Figure 1 [Bowden82], which we call a *ladder diagram.*[8]

The shaded shape points out a unique property of the homology-cohomology ladder, one that even many topologists seem unaware of, namely that the isomorphisms μ, μ^{-1} are *twisted*, that is, the kernel of the group at one end of a rung is mapped by μ (respectively, μ^{-1}) into the non-kernel elements of the group at the other end. [The isomorphisms μ, μ^{-1} are matrices containing the terms' $\mathbb{Z}_3$ coefficients.] This property was discovered by [Roth] in his proof of the correctness of Gabriel Kron's then controversial methods for analyzing electrical circuits [Bowden82], and turns out to have profound implications: the entirety of Maxwell's equations and their interrelationships can be expressed by a ladder with two rungs plus four terminating end-nodes [Bowden], and [Tonti] has - independently - shown similar relationships for electromagnetism and relativistic gravitational theory. Roth's twisted isomorphism (his term) thus reveals the deep structure of the concept of boundary, and shows that the complete story requires both homology and cohomology.

2.3 Generalizing the Twisted Isomorphism Hierarchy

Each level of a ladder hierarchy, as presented so far, is built entirely from entities (ie. sensors) from the level immediately underneath, leading to what we call a 'pancake' hierarchy. But this is an unnecessary limitation, from which we now generalize.

Let S_i be the set of sensors at δ-level i. Similarly, let G_i be the set of (sensors expressing the presence of) goals at ∂-level i. A pancake *meta* hierarchy of 2-actions can now be characterized by $S_i = S_{i-1} \times S_{i-1}$, where $\times$ is the cartesian product mediated by δ. Other, more general, hierarchical forms are now easily seen:

[8] Strictly speaking, ∂, δ, and μ/μ^{-1} should all be indexed by level: $\partial_\ell, \delta_\ell, \mu_\ell/\mu_\ell^{-1}$.

The left side of the ladder is the homology sequence generated by ∂ over the representation of actions as Clifford products cf. eqn. (3). The downward flow of decomposition of the structure into simpler pieces (ie. the crossing of successive boundaries) corresponds to the trickling down of goals to sub-goals described earlier.

The right side of the ladder is similarly the cohomology sequence generated by δ from sensory impressions, cf. eqn (4). The upward flow of composition of structure to form more complex structure corresponds to the effect of co-exclusion, up through which increasingly complex structure sensory impressions bubble.

The circles represent all the entities (Clifford algebra terms) at the particular level of complexity. The larger of the two circle halves holds those entities which will map to zero with the next hierarchical transition (∂ or δ) - the kernel of the group - as indicated by the pointed 'beak'.

The rungs of the ladder, besides denoting the location and content of hierarchy levels, also express the existence of isomorphisms (μ, μ^{-1}) between the structures at either end of a given rung. The shaded portion, which can be seen to repeat in both directions, expresses the commutation relations that obtain.

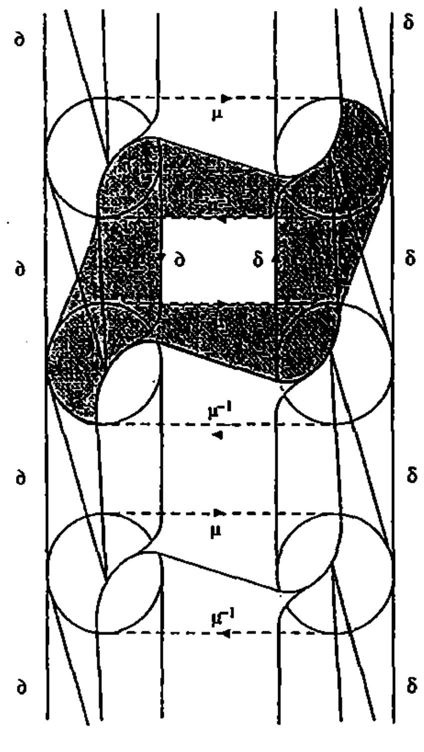


Figure 1: Ladder diagram, illustrating homology-cohomology relationships.

- $S_i = S_j \times S_k$, $j, k < i$, yielding non-pancake meta hierarchies; and of course the product may be over >2 levels. Aside from this, however, the semantics is roughly as before;
- $G \times G$, yielding a purely goal-based *icarian* hierarchy, roughly similar to a function-composition hierarchy;
- $S \times G$, yielding a combined abstraction over the underlying ladder level(s) that we call a *morphic* hierarchy.

Figure 2 illustrates the latter two, and we note that the morphic level in (a) may in principle 'cross' levels more radically, eg. as (b) does. We call these generalized forms *ortho-hierarchies*.

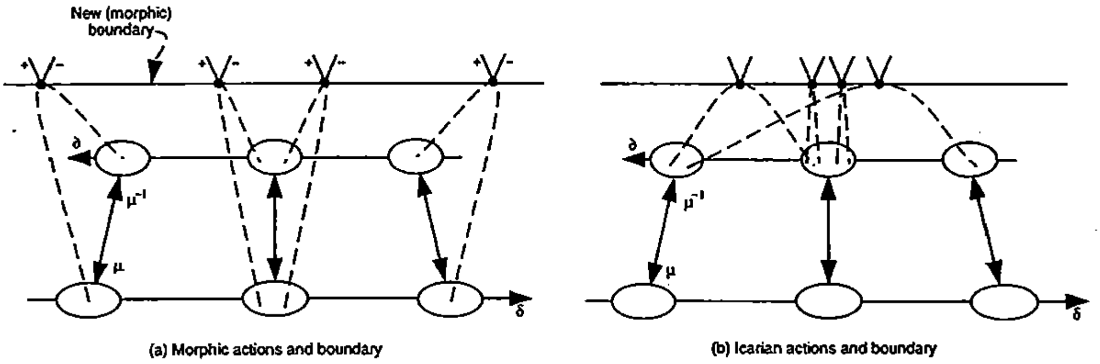


Figure 2: Morphic vs. icarian hierarchies.

Icarian actions provide a means for a computation to express, self-reflectively, the *way* it carries out its goals. Morphic actions provide a means for a computation to express, self-reflectively, the relationship between S and G that is otherwise buried in μ, μ^{-1}. In the following, we will use the term 'meta' to denote all three types of abstraction.

3 Meta-sensor ambiguity and its resolution

At this point, we have seen two informational mechanisms for emergence - co-occurrence and co-exclusion, a mechanism for constructing the latter from the former, and a mathematical framework that describes these and their composition into a very general hierarchical form. Still missing, though, is the mechanism that propagates information up and down this hierarchy.

The key issue here is that the algebraic meta-sensor symbol $s_1 s_2$ codes two bits of information - the orientations of s_1 and s_2 - into one bit - the orientation of $s_1 s_2$. Another way to put this is that the algebraic symbol does not distinguish between the duals: $s_1 + s_2 \leftrightarrow \tilde{s}_1 + \tilde{s}_2$ versus $\tilde{s}_1 + s_2 \leftrightarrow s_1 + \tilde{s}_2$.

Looking more closely at this, there are the following possibilities for a state-propagation mechanism. Ie. assuming that we begin in the state $s_1 + s_2$, flip the orientation of $s_1 s_2$

1. When both s_1 and s_2 flip.

   Here, we encode the two states $s_1 + s_2$ and $\tilde{s}_1 + \tilde{s}_2$ as the two possible orientations of $s_1 s_2$. Unfortunately, there is a problem: the two dual states $\tilde{s}_1 + s_2$ and $s_1 + \tilde{s}_2$ are not represented at all in the meta-sensor state, and if one of these latter states obtains, then the meta-sensor's state is undefined.

   If we therefore insist on a distinct meta-sensor for each of the duals, we give up any attempt at state compression (abstraction). We also get the problem of the very existence of the dual meta-sensor/meta-action...it may not even exist yet, and the changing of a single sensor is not enough to trigger its discovery (cf. the event window mechanism). This issue revolves around the fact that mathematically, all components of the space (here, Clifford algebra terms) are always implicitly present when needed, but this is not necessarily the case when, as we intend, they directly represent physical entities.

2. When one of s_1 or s_2 flips.

   In this encoding, one orientation of $s_1 s_2$ indicates one dual, and the other orientation the other dual. However, this does not distinguish between the two states within a given dual (Exactly which state are we in? In which direction did we rotate?!), and $s_1 s_2$ doesn't flip at all if s_1 and s_2 flip simultaneously. In effect, we double the rate at which the meta-sensor flips to (partially) compensate for the fact that two bits simply cannot be encoded in one bit.

   This alternative (dubbed 'symmetric')succeeds at compressing state only because there are exactly two duals (which by definition exclude each other).

3. According to the orientation of s_1.

   Like the preceding alternative, this accepts ambiguity in the phase of the current and resulting states, in this case distinguishing one half of the $s_1 \times s_2$ plane, but not which quadrant. Mapping s_2 has a conjugal effect. Allowing both maps (to distinct meta-sensors) removes the ambiguity at the price of losing abstraction.

⋆ Note that we have only treated 2-actions. An action of arity n possesses $\frac{n}{2}$ duals, so the above issues compound as arity increases. We will encounter this in the next section.

⋆ Note also that the existence of duals creates a *naming* problem, in that, on the one hand, all the duals of a given action should have a common name to reflect this familial relationship, while on the other they are all distinct from each other. Thus a given dual's name must be a 2-tuple of the form (actionId,dualId), where actionId is a (commutative) hash of (the names of) all the action's constituent sensors (both polarities), whereas dualId need only be locally unique.

The above list is couched in terms of the bubbling of state change up the δ-hierarchy. A similar analysis can be carried out from the point of view of goals trickling down the ∂-hierarchy, in which case the question is: given the goal $s_1 s_2 \rightarrow \widetilde{s_1 s_2}$, which of the possible subgoals $s_1 \rightarrow \tilde{s}_1$, $s_2 \rightarrow \tilde{s}_2$, or both, should be issued, and when should they be retracted?

From either point of view, the second of the above alternatives seems preferable, for the following reasons:

- The phase ambiguity can be seen as a nifty way to model non-deterministic outcomes (true also of the third alternative, but not as symmetric).
- The second alternative satisfies the identity $s_1 s_2 (s_1 s_2 + s_1 + s_2) s_2 s_1 = s_1 s_2 + \tilde{s}_1 + \tilde{s}_2$, and hence preserves the semantics we arrived at in the preceding section.
- The third alternative has the effect that higher-level abstractions continue to ape primitive-level sensors ad infinitum, rather than the new exclusions they purport to reflect.
- The third alternative introduces a new problem: how to choose which of the two constituent sensors is to be mapped to the meta-sensor?

Unfortunately, the second alternative only works for arity 2, but Nature's well-known affinity for symmetry should perhaps not be denied. Therefore we accept this hint and will try to solve our problems with arity 2 co-exclusions only. Nevertheless, since the choice of propagation model should have predictive consequences, this is an issue that can be resolved empirically (and we claim a certain amount of empirical support for our choice, as will become apparent).

4 The Bit Bang and quaternions

In this section we present a Big Bang scenario, but where, in contrast to the usual version, the expansion is in terms of information. This information is the result of making distinctions, and the maker is 'the universe' in the guise of an initial ***Void***. In using the latter term, we intend no particular a priori interpretation, whether physical, logical, or metaphysical. This said, interpreting it in the present context as the vacuum is natural.

Regarding the making of distinctions, and in line with our development thus far, we will apply the distinction 'co-occur vs. exclude', which pair has the distinction-defining property that the two aspects necessarily exclude each other. In that the very utterance of one half of a distinction implies the other, they become conceptually co-occurrant in yin-yang fashion. This fact is in turn captured by the co-exclusion inference, which simultaneously introduces the hierarchical moment analyzed in §2.3.

Before presenting our *Bit Bang*, however, it is appropriate to motivate its relevance to our present endeavor. There are two aspects, the first being to demonstrate the emergence of space in the form of quaternions. Equally important, however, is the more theoretical problem of grounding the endeavor as whole. The point here is that in a hierarchical theory, such as the one we are presenting, there are two components: the entities that populate the various levels (eg. quaternions) and the mechanism of the hierarchy itself, ie. the mechanism by which the hierarchy is constructed.

The latter is responsible for the basic properties of the entities, and *these properties are by definition the same at every level.* We call this property of a hierarchical theory *level independence*, and it is both the blessing and the burden of any hierarchical theory. In the case at hand, the basic properties are those of co-occurrences and (co-)exclusions, and derivations and implications hereof. Level independence thus implies that the phase web's hierarchy should be able to give a reasonable account of its creation ab initio, thereby *grounding* the entire construction. We interpret the ab initio construction as as an informational Big Bang, which information is the product of the distinctions afforded by co-occurrence (cf. the Coin demonstration) and co-exclusion (cf. the Block demonstration).

On the next page, then, is our Bit Bang. We divide it into a series of steps, where each step is intended to follow inevitably from the preceding one. One alternative branch is (as noted) from Step 0 to Step 1, where one could use $0 = \tilde{0}$ instead of $0 = 0 + 0$, yielding $\tilde{\mathit{1}}$ at Step 1 and thereafter reversing the logic of Step 2 to yield $\mathit{1}$; this branch thus rejoins the one given at the end of Step 2. In Step 1 one could also ask $\mathit{1}+0$, but this also yields $\tilde{\mathit{1}}$ Finally, Steps 2, 3, and 4 each 'close' a logical level.

Regarding the meta-physical language: what we are trying to convey cries out for verbal interpretation, and without such language the mathematical expression is more arcane than expressive, expecially in the beginning.

Step	*Symbolically*	*Commentary*
0	*Void* = 0	That about which nothing can be said. Even naming it implies the existence of something that is not *Void*. But *Void* is everything and nothing, paradoxically simultaneously thinkable and unthinkable. Physically, *Void* is (presumably) the vacuum. Mathematically, we might attempt $0 = 0 + 0 = \tilde{0} = \tilde{0} + \tilde{0} = 0 + \tilde{0}$ but even this reifies distinctions we are forbidden: '+' implies 'parts', and '$\sim$' implies 'non-*Void*'. But the universe undeniably exists! So from *Void* there must be a step. Suppose it was $0 = 0 + 0$, ie. the parts are as the whole (starting with $0 = \tilde{0}$ yields $\tilde{I}$, thence I). Denote this distinction by the symbol...
1	I_0	which means 'the same as'. But, having now admitted 'parts', we must ask, What is $I_0 + I_0$? [We now invoke $\mathbb{Z}_3$ because: (1) $\mathbb{Z}_0$ is not open to extension, (2) $\mathbb{Z}_2$ doesn't distinguish *Void* from 'opposite' so (3) $\mathbb{Z}_3$ is the first possibility. Since co-occurrence together with exclusion exhaust/fulfill *Void* (see next step), it appears that $\mathbb{Z}_3$ is sufficient to all future expansion.] Presuming then $\mathbb{Z}_3$, the answer is
2	$\tilde{I}_0$	that is, I_0 is not the same as its parts ... $\tilde{I}_0$ means 'the opposite of', and
	$I_0 + \tilde{I}_0$	means that the parts 'the same as' and 'the opposite of' $\equiv$ *Void*. Ie. the marriage of sameness and oppositeness exhausts/fulfills *Void*. Denote this latter distinction, which is *new*, by...
3	I_1	$I_0 + \tilde{I}_0 \rightarrow I_1$ is an "arity 1" co-exclusion, δ_0. [Symbols with subscript $= 1$ are in effect 'discrete variables in $\mathbb{Z}_3$'.] Now that we have both *sameness* (co-occurrence) and *oppositeness* (exclusion), we can ask, What is $I_0 + \tilde{I}_0 + I_0 + \tilde{I}_0 = I_1 + \tilde{I}_1$? This distinction is a true (arity ≥ 2) co-exclusion. [This step, and similar ones later, assumes that the *Void* can/will continue to produce new step-1 instances as needed. Logically, this is unproblematic; physically, it assumes the same vacuum activity as produced the first instance.] The result of the co-exclusion of I_1 and $\tilde{I}_1$ is
4	I_2	and, via step 2, $\tilde{I}_2$ follows. Note that $(I_2)^2 = (\tilde{I}_1)^2 = -1$, cf. §2.1.

Clearly, we could continue this listing of distinctions ad infinitum, but we choose to end it here, since step 4 has yielded $s_1 s_2$, which is the basic quaternion building block.

Via the mappings

$$I_0 \mapsto 1, \quad \tilde{I}_0 \mapsto -1$$
$$I_1 \mapsto s, \quad \tilde{I}_1 \mapsto \tilde{s}$$
$$I_2 \mapsto s_1 s_2 \mapsto e_1$$
$$I_2 \mapsto s_2 s_3 \mapsto e_2$$
$$I_2 \mapsto s_3 s_1 \mapsto e_3$$

it is easy to verify the defining quaternion relations

$$e_i^2 = -1$$
$$e_i e_j = -e_j e_i, \quad i \neq j$$
$$e_1 e_2 = e_3, \quad e_2 e_3 = e_1, \quad e_3 e_1 = e_2$$

whence we have redeemed the promissory note contained in the title of this paper.

Notice however that we have only witnessed the emergence of *local* "3-D-ness". We are *not* claiming (nor do we wish to claim) that this 3-D-ness is a global Newtonian space with unique origin, *nor* even relativized multiple ditto. Rather, 3-D-ness is a property of co-exclusion-derived objects with sufficient information-carrying capacity ("complexity"). The globality of 3-D-ness can only be achieved by the need for consistency between objects sharing a given distinction (sensor), and a change in the state of a given distinction must therefore propagate through the structure. Thus it appears that our construction is entirely consistent with, although conceptually 'under' or 'prior to', general relativity in these respects. Moreover, since some distinctions lie below the level at which global 3-D-ness emerges, changes in these can appear to propagate more rapidly, since they are not constrained by the higher level structures (which will nevertheless always behave consistently vis à vis such changes). This is the phase web's way of reconciling the locality conflict between relativity theory and quantum mechanics.

Given that we now have the three quaternion operators and the (local) 3-D spatial properties they define, it is natural to ask if the hierarchical buildup also can produce the 3-D objects we expect to find in such a space. We now address this question.

Our everyday experience tells us that three spatial dimensions can hold three-dimensional objects, so we should expect the extension to be straightforward. And it is, since $\delta(s_i s_j + s_j s_k + s_k s_i) = s_i s_j s_k$ and $s_i s_j s_k$ is an oriented volume (although it will develop that this is not *quite* right). [We postpone the issue of finding some mass to fill this volume.] Notice by the way that this formulation requires an arity-3 co-exclusion.

The next issue is how to propagate state up to this new, volumetric, entity. The problem is the same as before, except worse: instead of needing to encode two bits into one, we now must encode three into one, that is, reflect eight possibilites in two. Although we will eventually arrive at a similar solution as before (ie. arity 2), the details are instructive.

The table below lists the eight possibilities (viewing the three rightmost columns as binary numbers, the first column's numbering is the decimal equivalent):

	s_1s_2	s_2s_3	s_3s_1
7	1	1	1
6	1	1	$\tilde{1}$
5	1	$\tilde{1}$	1
4	1	$\tilde{1}$	$\tilde{1}$
3	$\tilde{1}$	1	1
2	$\tilde{1}$	1	$\tilde{1}$
1	$\tilde{1}$	$\tilde{1}$	1
0	$\tilde{1}$	$\tilde{1}$	$\tilde{1}$

The pairs **7,0**, **6,1**, **5,2**, **4,3** are co-exclusions, and are distributed symmetrically about the horizontal line. If we simply co-exclude these amongst themselves, we will get even more (six, to be exact) so this approach diverges. Rather, if we are to use an encoding similar to that of the earlier 2-coex case, we must look at little more closely at the dynamics of these entities. One could say that for these four 3-co-exclusions, the dynamics is that all three bits flip. What, then, if only one or two flip?

One change at a time yields so-called Grey-coded sequences, and the connectivity of the transitions is captured by a unit cube, each of whose vertices is labelled by one of the above states. That is, no compression of states occurs. We conclude therefore that this kind of distinction is not useful (nor is it a co-exclusion, so it's not really a valid distinction anyway).

In the case where two meta-sensors flip (and one thus remains constant), it turns out that there are two disjoint families, **0**, **3**, **5**, **6** and **1**, **2**, **4**, **7**, each defining a tetrahedron, ie. a plane plus a point outside of that plane, and hence 3-D orientation. See Figure 3.

Observe now the following:

- Half of each family is above/below the line, and the members of the halves pair complementarily, so above/below the line each contains all four *rotation* states around s_1s_2;
- Each family as well expresses the four possible *rotations* around s_1s_2, so the same information regarding the two bits that are flipping is available both to each family and above/below the line.

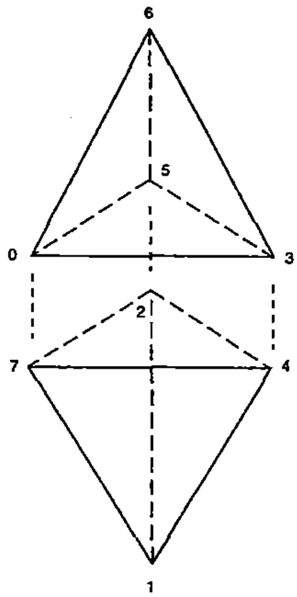


Figure 3: The eight states, via 2-flip transitions (denoted by the edges), form two tetrahedra of opposite handedness (parity).

- A transition across the line (ie. between two 3-co-exclusive states) is a *reflection* (via a factor of -1, cf. $ab(a+b)ba = -(a+b)$);

There are three different possible meta-sensors, depending on which of s_1s_2, s_2s_3, s_3s_1 we choose to be the first column. To make use of our symmetric arity-2 meta-sensor construction/restriction, we choose these three meta-sensors to be the three co-exclusions (circularly) of s_is_j with s_k, ie. $\delta(s_is_j + s_k)$, written $s_is_j|s_k$. That such a meta-sensor form in fact describes $s_is_js_k$ is guaranteed by the fact that

$$\partial(s_1s_2s_3) = -[\partial(s_1s_2)s_3 + \partial(s_2s_3)s_1 + \partial(s_3s_1)s_2]$$

∂ on the righthand side can be interpreted as the corresponding events (=sensor changes).

Recalling (cf. §3) that a symmetric 2-meta-sensor flips when one, but not both, of its two constituents flips, the following table shows what happens to the three $s_is_j|s_k$-meta-sensors when, respectively, one, two, or all three constituents flip. *Notation*: $\times$ means a flip, a $-$ means no flip.

What flips	s_is_j	$\vert$	s_k	s_js_k	$\vert$	s_i	s_ks_i	$\vert$	s_j	Total
s_i	$\times$		$-$	$-$		$\times$	$\times$		$-$	3
s_i, s_j	$-$		$-$	$\times$		$\times$	$\times$		$\times$	0
s_i, s_j, s_k	$-$		$\times$	$-$		$\times$	$-$		$\times$	3
		P			C			I		

Since a symmetric meta-sensor only flips when *one* of its constituents flips, the only changes that count are those that pair a – with a ×, as reflected in the rightmost column. Thus a flip of one of the three sensors causes all three meta-sensors to flip, whereas a flip of two causes none, and when all three base-level constitutents flip, so do all three meta-sensors. We now examine each case more closely; the bottom row assigns the names P,C,I respectively to the three mixed-level meta-sensors. (It is useful in thinking about this to have a picture of a little 3-D coordinate system in mind.)

Only s_i flips. This causes a *reflection* of of both the s_i, s_j, s_k and PCI coordinate systems.

Both s_i and s_j flip. This causes a *rotation* in the s_i, s_j, s_k coordinate system, but no change in the PCI coordinate system (although the changes within P, C, and I are real enough).

All of s_i, s_j, s_k flip. This causes a *reflection* in the both the s_i, s_j, s_k and PCI coordinate systems.

Reminding ourselves of the CPT symmetry, this is *exactly* what should happen had we denoted the P meta-sensor as *parity*, the C meta-sensor as *charge*, and the I meta-sensor as *isospin* (which denotes the projection of the charge in one of three spatial directions).

In independent support of this identification, we tentatively offer the following. There are six quarks, occurring in families of two each, these two differing most critically in their charge: $+\frac{2}{3}$ vs. $-\frac{1}{3}$. In the mathematical formulation, $s_1s_2$and $s_2s_3|s_3s_1$ are operationally equivalent, but in the actual realization, the latter is a distinct co-exclusion whose result just happens to have the same effect as s_1s_2 but 180° out of phase. Because s_1s_2 is half the 'size' of $s_2s_3|s_3s_1$, we assign it charge $-\frac{1}{3}$ and the latter $+\frac{2}{3}$. The nicely logical way this works out together with the way changes in the three basis sensors $\{s_1, s_2, s_3\}$ are coupled across the PCI meta-sensors in CPT-like fashion as just described, argue for identifying these three meta-sensors with parity, charge, and isospin.[9]

We have been tempted to speculate in such matters in order to argue for our quaternion construction, and hasten to add that there are many details of the above quark structure that must be checked. In any event, it is important for the reader to understand that the only degree of freedom in this little game lies in *how* to arrange the pieces, ie. the number and arity of possible co-exclusions and their mappings to corresponding meta-sensor states. *All* co-exclusions denote distinctions (ie. bits) that the universe can and will make, which distinctions people denote by various quantum numbers (generally elements of $\mathbb{Z}_3$) eg. spin, parity, and charge. Hence, the combinatorial structure *must* provide every particle (known or otherwise) with a unique and consistent placement in that structure - one misfit means that the whole idea dies. In all cases, spin, quaternions and local 3-D-ness, parity,

[9]Having opened Pandora's box here, we speculate that *mass* is proportional to the number of bits (distinctions) enclosed by a given co-exclusion envelope, but a glance at the quarks' empirical values shows that there is more to the story.

charge, and isospin are all clearly seen to be both distributed and emergent, and all are properties of the object $s_1 s_2 s_3$.

There is one final categorization of distinctions we must mention, namely that described by The Combinatorial Hierarchy (CH) [Bastin&Kilmister, Parker-Rhodes]. This hierarchy is traditionally constructed in $\mathbb{Z}_2$, but there is general agreement that it and the $\mathbb{Z}_3$ Bit Bang presented here are in some sense isomorphic. Howsoever, the key point is to examine the number of *discriminately closed subsets* (dcs's), that is, subsets that close under the discrimination operation (in our case, exclusion; in the CH's, exclusive-or). These are (cf. Figure 4)

$$\{s\}$$
$$\{s_1, s_2, \{s_1, s_2, s_1 s_2\}\},$$
$$\{s_1, s_2, s_3, \{s_1, s_2, s_1 s_2\}, \{s_2, s_3, s_2 s_3\}, \{s_3, s_1, s_3 s_1\}, s_1 s_2 s_3\}$$
$$\ldots$$

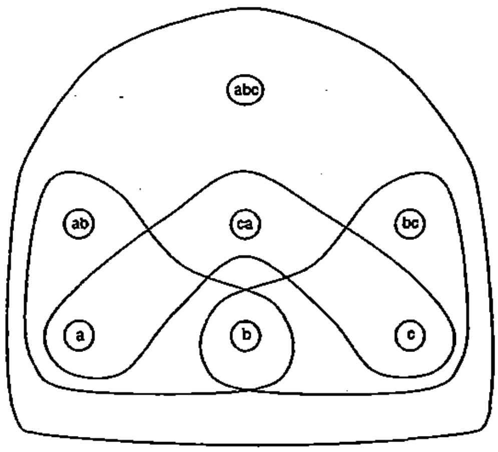


Figure 4: Combinatorial Hierarchy categories (dcs's) in $\mathbb{Z}_3$.

The key numbers are the successive sums of the dcs cardinalities: $(1), 3, 7, 127, 2^{127} - 1 \longrightarrow 3, 10, 137, 1.7 \times 10^{38}$, ie. column c in the table below.

(*a*) level	(*b*) # symbols per level	(*c*) cumulative $\sum(b)$	(*d*) map dim.	(*e*) # of map elements	(*f*) comment
0	(1)	(1)	(1)	(1×1)	
1	3	3	4	$4 \times 4 = 16$	$16 > 7$
2	7	10	16	$16 \times 16 = 256$	$256 > 127$
3	127	137	256	$256 \times 256 = 65536$	$65536 < 2^{127} - 1$
4	$2^{127} - 1$	$2^{127} + 136$	$(256)^2$		cut-off reached

Column (b) is simply the full number of ways a number of entities (symbols) can be combined - 1,2,3... at a time, which is $\sum_{p=1}^{n}\binom{n}{p} = 2^n - 1$. This sequence thus counts the number of symbols that can be formed from some given set of symbols by aggregation. A second sequence, column (d), is related to the number of symbols from column (b) which can via discrimination produce the remaining ones at the next level.

Especially the last two numbers in column c are thought provoking: more detailed combinatorial calculations yield a corrected value of the inverse fine structure constant very near the experimental value (137.0359 674 vs. observed 137.0359 895(61)), and similarly for the ratio of the electromagnetic and gravitational forces ($2^{127} + 136 = 1.69331 \times 10^{38}$ vs. observed $1.69358(21) \times 10^{38}$), respectively. Interestingly, the sequence in column b cuts off after the fourth step, since a symbol-basis of 65536 cannot span a space with $2^{127} + 136$ elements. See [Noyes] for these and a number of other physical constants calculated on this purely combinatorial basis. Note also that the dcs's correspond to meta/morphic constructions (cf. §2.3) restricted to closure.

5 Summary and conclusions

We have described a truly distributed model of computation - the phase web - based on the distinction between co-occurrence and mutual exclusion of both states and events. This model, by virtue of its acceptance of true concurrency, exceeds Turing's model of computation (which conclusion, while not widely appreciated, is not controversial in the computer science community). The importance of a *computational* model, in contrast to so many other kinds, is that it provides explicit mechanism, and we argued for the utility of mechanism as a tool for reasoning, not least in the context of 20^{th} century physical theory. More concretely, without §3's search for a mechanism for propagating state through the hierarchy, §4's quaternion result would have been elusive, and perhaps impossible.

The fundamental hierarchy-building operation of co-exclusion - which expresses emergent phenomena naturally - turns out to be nicely modelled by Clifford algebra's product, which algebra can thereafter be emplaced in the topological context of the twisted isomorphism between homology and cohomology. The coboundary operator δ was seen to correspond to, precisely, co-exclusion; and the boundary operator ∂ to action. Thus the hierarchical moment implicit in the co-exclusion operation became mathematically explicit.

The fact that each level of the hierarchy is built via the same operation leads to the concept of the level independence of phenomena. Level independence is what gives power and scope to hierarchical theories, but also carries with it the complementary burden of showing that it truly does apply to any level of description, or, if you will, empirical fact.

Turning this around, *if* we are to theorize meaningfully about (say) consciousness - which we believe our model can accomodate - we should have some reason to believe that our theoretical framework is grounded in reality.

To establish this, we modelled the cosmological Big Bang as a process of informational expansion deriving from the progressive compounding of distinctions, each distinction (co-exclusion) expressing one bit of information. Our demonstration in this paper of the emergence from this process of local 3-D-ness in the form of quaternions, besides its intrinsic interest, thus also allows us to discuss more complex phenomena and systems with rather greater confidence.

Relative to the quaternions themselves, we saw that the local 3-D space they define requires prior structure possessing sufficient information-carrying capacity to express the distinctions associated with 3-D-ness, of which the crucial one is parity. We saw that two other distinctions, intertwined with parity, appeared at the same time, which, inspired by the CPT theorem, we tentatively identified with charge and isospin. This broaching of the topic of particle structure gives a another way to test the validity of the points of view being advanced in this paper.

The extension of local 3-D-ness to 3+1 spacetime remains, and the path to be followed seems clear, although undoubtedly rocky.

Acknowledgements.

The basic structure of the Bit Bang was inspired by The Combinatorial Hierarchy of [Parker-Rhodes] and [Bastin&Kilmister], which in turn is based on a construction of the integers originally due to Conway. The Coin and Block demonstrations are reproduced with IEEE's permission from [Manthey94]. Special thanks to Rainer Zimmerman and Achim Müller for their gracious hosting of the *Natura Naturans* '97 workshop in Bielefeldt, where a preliminary version of this work was presented.

References

Bastin, T. and Kilmister, C.W. *Combinatorial Physics*. World Scientific, 1995. ISBN 981-02-2212-2.

Bastin, T. and Kilmister, C.W. "The Combinatorial Hierarchy and Discrete Physics". Int'l J. of General Systems, special issue on physical theories from information (in press).

Bowden, K. "Physical Computation and Parallelism (Constructive Post-Modern Physics)". Int'l J. of General Systems, special issue on physical theories from information (in press).

Bowden, K. *Homological Structure of Optimal Systems*. PhD Thesis, Department of Control Engineering, Sheffield University UK. 1982.

Feynman, R. *The Character of Physical Law*. British Broadcasting Corp. 1965.

Hestenes, D. and Sobczyk, G. *From Clifford Algebra to Geometric Calculus*. Reidel, 1989.

Hestenes, D. *New Foundations for Classical Mechanics*. Reidel, 1986. The first 40 pages contain a very nice, historical introduction to the vector concept and Clifford algebras.

Manthey, M. "Synchronization: The Mechanism of Conservation Laws". Physics Essays (5)2, 1992.

Manthey, M. "Toward an Information Mechanics". Proceedings of the 3rd IEEE Workshop on Physics and Computation; D. Matzke, Ed. Dallas, November 1994. ISBN 0-8186-6715X.

Noyes, H.P. SLAC-PUB-95-7017 and SLAC-PUB-7205, pp-36-38.

Parker-Rhodes, F. *Theory of Indistinguishables - A Search for Explanatory Principles Below the Level of Physics*. Reidel. 1981.

Penrose, R. *The Emperor's New Mind*. Oxford University Press, 1989. ISBN 0-19-851973-7.

Roth, J.P. "An Application of Algebraic Topology to Numerical Analysis: On the Existence of a Solution to the Network Problem". Proc. US National Academy of Science, v.45, 1955.

Tonti, E. "On the formal structure of the relativistic gravitational theory". Accademia Nazionale Dei Lincei, Rendiconti della classe di Scienze fisiche, matematiche e naturali. Serie VIII, vol. LVI, fasc. 2 - Feb. 1974. (In english.)

www. Various phase web and Topsy publications, including (soon) code distribution, are available via *www.cs.auc.dk/topsy*.

The Categorial Approach to the Part-Whole Relation: Mereological Extra-Level Emergence as the Emergence of new Limits

Diagrams and Figures by Jens WALDECK

Peter EISENHARDT; Dan KURTH; Jens WALDECK

Institut für Geschichte der Naturwissenschaften
Johann Wolfgang Goethe Universität
D-60054 Frankfurt am Main, Germany
e-mail: Eisenhardt@em.uni-frankfurt.de; FSCTJensW@aol.com

In several previous papers[i] we have argued for a formal access to a required solution of all the intricacies of the still not thoroughly formulated theory of emergence. In these we always decisively favoured an approach based on category theory. Therefore we had been agreeably surprised when we eventually found out that our obstinate conviction (that the appropriate formalism to describe emergence might be of a categorial kind) was somewhat corroborated by an earlier paper of A.C.EHRESMANN and J.-P.VANBREMEERSCH about "Hierarchical Evolutive Systems"[ii] in which EHRESMANN/VANBREMEERSCH also proposed a framework in terms of category theory to explain the emergence and complexification of systems in general up to hierarchical ones. But then there are - despite some similarity with our formal access to emergence - some significant differences in their otherwise very interesting outline, as well in the particular scope of their model as in the underlying ideas of the appropriate applicability of such mathematical models.

The first of our objections is directed to what one might call a kind of a pythagorean mistake in the sense of "giving the number of the organism"[iii]. By ascribing a proper category to any particular individual organism the authors treat 'complexity' somehow as an intrinsic attribute of category theory, but - despite a possible fit of category theory for the mathematical treatment of emergence and complexity - this would hardly be the case. Organisms have no numbers notwithstanding that one can describe them quantitatively. So as fervently we agree that 'complexity' is a property of natural objects or systems which is describable in terms of category theory as strongly we deny it to be a 'categorial attribute' of these entities. Individual entities, systems or organism are not categories themselves - rather they must be described by means of category theory in such a way that the universe of discourse consists in structures or complexity classes. The appropriate biological counterpart of such an elementary complexity class is - despite all its taxonomical ambiguity - covered by the concept of 'species'.

Furthermore EHRESMAN and VANBREMEERSCH do not seriously enough tackle the question 'Why are there levels of nature?', i.e. they cannot describe the emergence of levels appropriately because - in our view - they simply insufficientely specify the emergentic act. I.e. due to the mentioned 'pythagorean mistake' they on the one hand misplace the emergentic event at the individual rather than the structural domain, and then consequently - but even worse - they have to restrict emergence to a mere special case of it, namely the case which find its most characteristic expression in sexual reproduction. Generally speaking EHRESMANN / VANBREMEERSCH restrict their concept of emergence or the evolution of increasingly more complex systems to a well known - and admittedly important - case namely a generalized version of the merger of two preceedingly autonomous or independent less complex entities into a succeeding new one of an increased complexity, i.e. an entity on a higher hierarchical level. This approach then should cover most of what is required for the description of an emergentic increase of complexity in the important - yet still quite special - case of the part - whole integration, i.e. the integration of two less complex parts into a new more complex whole. By tackling this original problem of even the earliest attempts of the philosophy of emergence with the appropriate means of category theory EHRESMANN / VANBREMEERSCH made at least a decisive step in the direction of a general solution of most of the problems entangled with the phenomenon of emergence.

Let us get to the core of the matter. The pythagorean ansatz says that a new entity or property emerges out of other ones by limit formation in the framework of category theory. A limit for a diagram (consisting of objects and arrows) is the same thing as a specific cone over the diagram.

If one takes a diagram with two vertices and without edges (two objects A and B; no arrows) the cone in the sense of EHRESMANN and VANBREMEERSCH (actually it is a cocone) over this diagram has one object P and two arrows p and q :

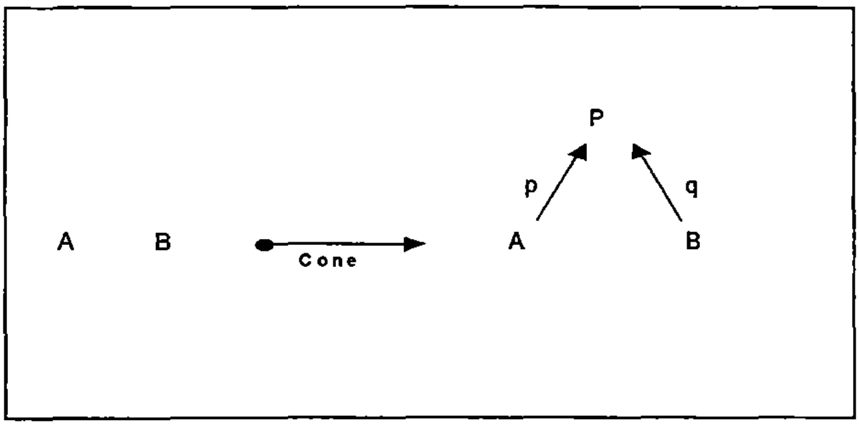


Fig. 1: Cocone of Two Objects

Let's stick to the example. The cone shown above becomes a limit iff it is a product diagram for A and B.[iv] A product diagram for our cone (A, P, B) has an additional property, namely an object Z with

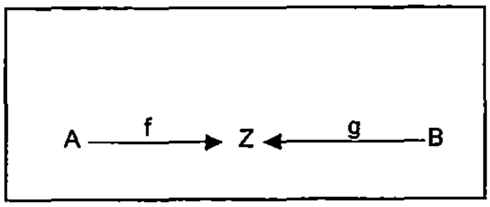


Fig. 2: Product Diagramm of the Cocone

and

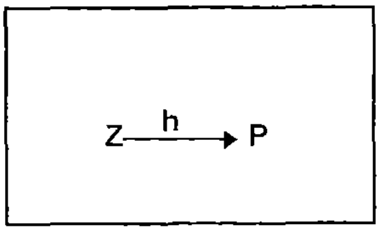


Fig. 3: From Product to Cocone

To simplify the representation we stretch the cone A, P, B and bend the limit property A, Z, B. The following amalgamated commutative diagram is the limit of A and B:

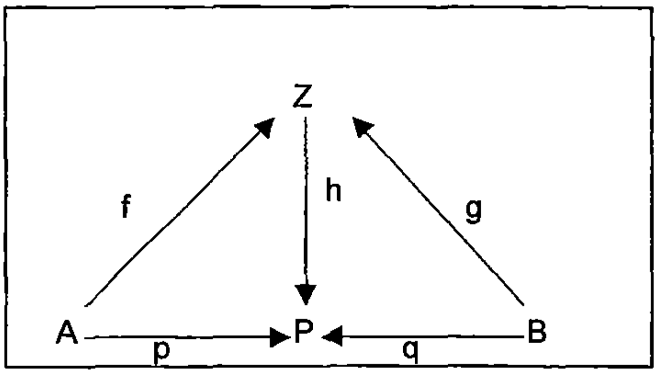


Fig. 4: Product as a Limit

If the topological product of two line segments A' and B'results in a rectangle A' × B' = P' one has in category theory the extra arrow $\downarrow_{h'}$ from Z' to P'guaranteing the universality of the transformation.[v]

EHRESMANN and VANBREMEERSCH make use of the limit-concept to describe the emergent tranformation from one level of complexity to a higher one.

> The idea is the following one: an object in a category has such a complex structure when itis composed of a family of more elementary objects, 'glued' together, the glueing depending on some specified links between the components.
> For instance, a word glues the letters it is formed of, in a given order, and the same letter may be repeated.[vi]

What does it mean for our example? The arrow $\downarrow_{h'}$: Z' → A' × B' guaranteing the universality of the product - formation and determining the product Z' and the arrows p' and q' "up to an isomorphism"[vii] "is said to *glue* the [direct product arrows p', q', ...r'] (the authors using their example - the number of the factors is unlimited)."[viii] So in our product diagram the factors (line segments) A' and B' with the (projection) arrows p' and q' give rise in a universal way (because of h') to the product P' (rectangle).

Naturally limit-formation is not restricted to a special product diagram. EHRESMANN and VANBREMEERSCH refer to any pattern of linked objects A_i, A_j,..., A_n in a category which shall be glued together by a limit. Let us have a look at the links staying with our example. The line segments A' and B' must be related by an arrow directly which represents their link 'before' they merge into a rectangle, this link being the cartesian coordinates in the plane.[ix] We have therefore $A'(x_m, y_m; x_n, y_n)$, $B'(x_m, y_m; x_n, y_n)$ or $A' \ {}_{c'}\!\rightarrow\ B'$. The complete diagram goes as follows (we bend the cone again and link A' and B'):

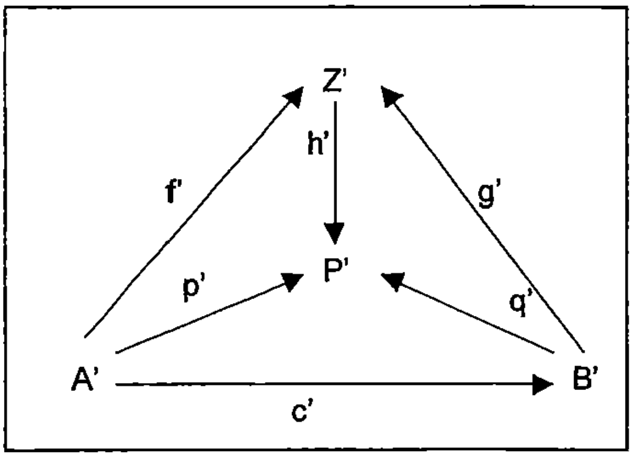


Fig. 5: Limit of Linked Objects

If we ignore the links (as we have done) the limit means only the *sum* of a family A_i, A_j, ... , A_l *of objects*[x] of level n-1 in contrast to the *complex object* of level n being more than the sum of its parts which emerges from the limit-formation of linked patterns (or 'arrowed' *objects* of level n-1). Without the latter no rectangle emerges but only two line segments arranged contingently.[xi] (Our authors think of more robust examples as molecules out of atoms or especially organisms out of cells. Freely adapted from EHRESMANN and VANBREMEERSCH the latter have to be described as patterns of linked objects glued together resulting in an organism via limit-formation.)

Four problems are mentioned:

1. "The emergence of properties, which is *measured* by an arrow obtained by comparing a complex object with the sum of its components..."
2. "The complexification of a system, thanks to the formation of new complex objects...Whence a construction of 'based' hierarchical systems..."
3. "The preservation of the identity of a complex component in a hierarchical evolutive system..."
4. "The formation of an evolutive system stepwise (...): new objects are absorbed, some components disappear, more complex components are formed, while others loose or preserve their organization..."[xii]

First of all let us come back to the pythagorean mistake mentioned above before we comment the four points.

This mistake is a very subtle one. On the one hand mathematical structures are used as tools or means to describe the scientific reality, on the other hand this reality seems to consist of those structures. Care is needed here. Certainly one gets no reliable information if one pictures the forms of the plants by laying little stones in a topological order on the floor asserting this to be the 'number of the rose'. But it makes sense to work out a precise topological model of let's say the growth and selection of plants and especially roses. In this way the transformation of a species (not individuals) will be described by means of topology but it is no topological property. (The sex of an organism is not equivalent to the genus of its topological structure.) Particulary complexity is no property of category theory but is describable in the framework of this theory, to be precise it is describable at the level of a *representation space* referring to the process of complexification. EHRESMANN and VANBREMEERSCH proceed straightforwardly:

> A complex system is modeled by a category whose objects 'are' its components, and whose arrows (or morphisms) are their interrelations.[xiii]

They propose a one-to-one mapping from a category onto a brain, an organism, or a molecule, for example. The objects 'are' (mapped onto) the neurons, the cells, and the atoms, whereas the arrows are the respective interactions. We consider that to be - in a somewhat exaggerated literal meaning, since it is accidentally related to the use of category theory - a category-mistake because the difference of 'object space' and representation space is leveled down.

Let us now comment on the four central points of the EHRESMANN/VANBREMEERSCH approach. The emergence of new properties is a consequence of the famous limit formation which we represented with a little help from geometry. The topological product of two line segments (point set intervals) gives a rectangle - in category theory one generalizes this procedure by the formation of a limit. But we can go on. The product of a line segment and a rectangle (surface) forms a solid

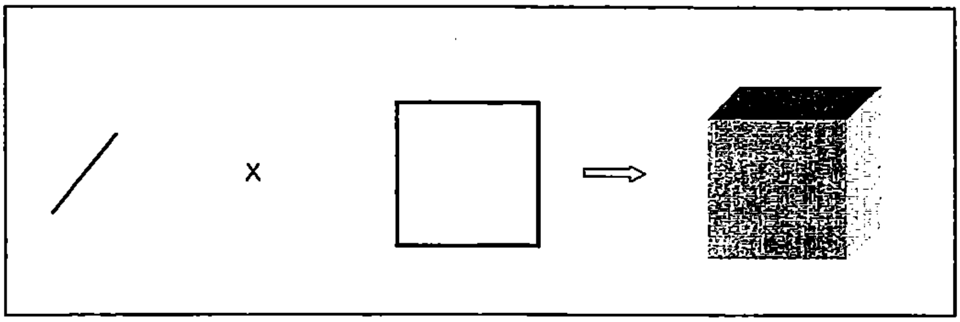


Fig. 6: Topological Product of a Line Segment and a Rectangle

the topological product of two circles $S_1 \times S_1$ gives a torus,

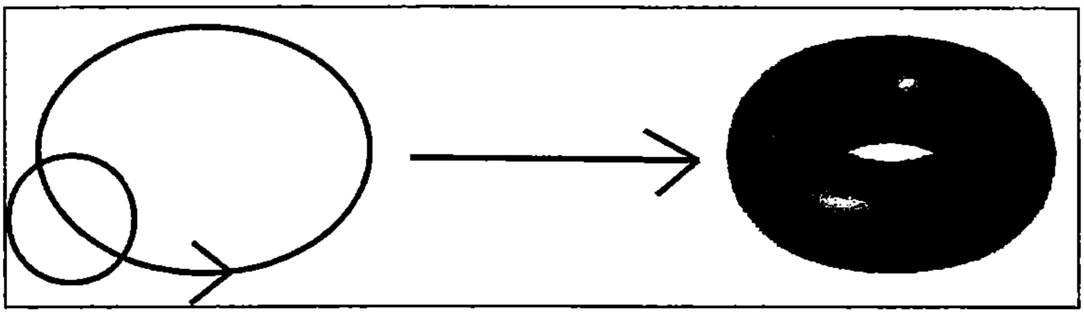


Fig. 7: Topological Product of Two Circles

three line segments form a cuboid too, a polyhedron of dimension 4 could be constructed out of two quadratic surfaces, two polyhedrons of dimension d and d' result in a polyhedron of dimension d + d' [xiv] etc. Here the idea of levels is introduced. A figure of level n (a rectangle for example) come up out of figures of level n-1 (line segments for example). Because one needs more coordinates to describe figures of higher dimensions the latter are in a sense more 'complex' than figures of lower levels.

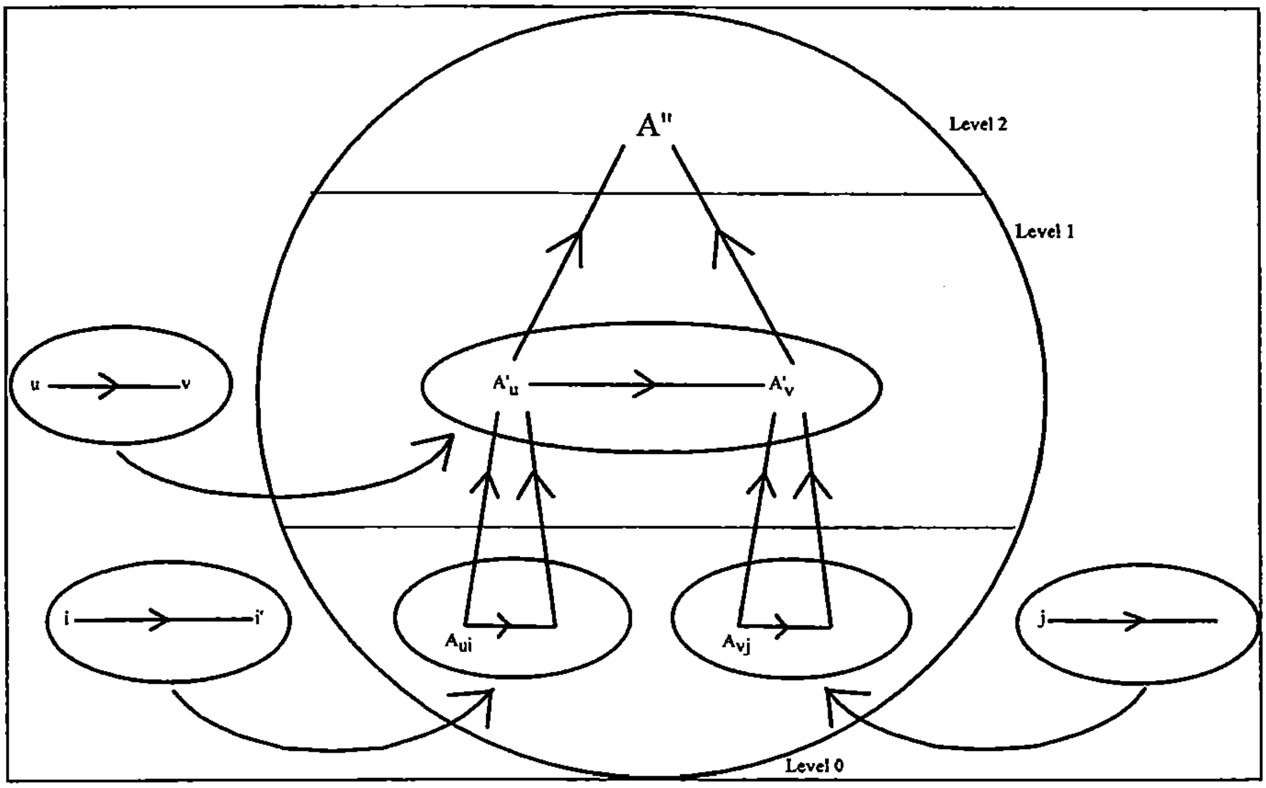


Fig. 8: Levels of Emergence as Iterated Limits (from Ehresmann/Vanbremeersch)

For EHRESMANN and VANBREMEERSCH an object of level n is the limit of a pattern of linked objects of level n-1. In that way they build a hierarchy of iterated limits. The molecules are the limits of interacting atoms (chemical bond for example), the organisms are limits of linked cells, and societies are limits of the pattern of the actions of the organisms (one can also 'study' the limit chain atoms → molecules → physical bodies → stars → galaxies, the authors remark.[xv] The objects can preserve their identity in the course of evolution though their parts are replaced, and more and more complex objects are formed in time (evolutive systems).[xvi]

Even our simple mathematical example (the product formation) shows a certain flaw of the theory of hierarchical evolutive systems. The theoretical surplus is in short supply because such an applied category theory imports only new names into an 'old-fashioned' categorial framework. A limit formation is nothing new and it is well-known that new spaces are formed from familiar ones by means of the (cartesian) product. So EHRESMANN and VANBREMEERSCH christen a familiar child "emergence" but gave no new theory and no new explanations. It is like introducing non-abelian groups as a theory of self-organization. They retell the whole story of the complexification of nature not telling us the deep structure or mechanism of the emergentic process.

So neither the mechanism of the morphological formation of new objects (our knotting) nor the 'uplifting' of new levels (or the developing of levels as such) are explained. Furthermore the ansatz suffers from depending on a sort of abstract sexual reproduction. Two (or more) entities (or patterns) are merged into another new entity. Several line segments produce a polyhedron, several atoms fuse to give rise to a molecule, and so on. In some respects this procedure is not false but it is liable to be misunderstood because of insufficient generality.

---

[i] EISENHARDT, Peter, Dan KURTH [1993], *Emergenz und Dynamik. Naturphilosophische Grundlagen einer Nichtstandard Topologie*, Cuxhaven Junghans-Verlag; EISENHARDT, Peter, KURTH, Dan, WALDECK, Jens, »Emergence as Antimorphic Action«, in: Philosophies, Proceedings of ANPA 17 (ed. Keith BOWDEN), September 1996; EISENHARDT, Peter, KURTH, Dan, WALDECK, Jens, » Emergence, Complexity and Integrative Levels «, in: Proceedings of ANPA 18, to appear

[ii] EHRESMANN, A.C.; VANBREMEERSCH, J.-P.[1987], »Hierarchical Evolutive Systems: A mathematical model for Complex systems«, Bulletin of Mathematical Biology 49 (1987) 13

[iii] Cf. ARISTOTELES: Metaphysics 1092 b 8f

[iv] MCLARTY, Colin [1992], *Elementary Categories, Elementary Toposes*. Oxford 1992: Clarendon Press, p.49.

[v] MACLANE, Saunders [1986], *Mathematics, Form and Function*. New York etc.1986: Springer-Verlag, p.393. The product diagram A $_p\rightarrow$ P $\leftarrow_q$ B is universal iff there is a unique arrow from Z to P. Replace Z by another product P' in the same category, namely A $_{p'}\rightarrow$ P' $\leftarrow_{q'}$ B' so that there is an arrow h' from P' to P with p'h'= p and q'h'= q:

$$\begin{array}{ccccc} & & P' & & \\ & {}_{p'}\nearrow & \downarrow_{h'} & \nwarrow_{q'} & \\ A & {}_{p}\rightarrow & P & \leftarrow_{q} & B \end{array}$$

The action of the operator h' on p' and q' gives the product immediately because of the commutativity of the diagram. So every other product P'''' transforms itself into a product again with the help of h'''' - h determines the product "up to an isomorphism", as MACLANE says. Consider that the arrows of the usual product formation reflect the so called surjective projections p: <x, y> → x and q: <x, y> → y with arrows P → A and P → B instead of our inverted arrows giving a commutative diagram too.

[vi] EHRESMANN, A.C.; VANBREMEERSCH, J.-P.[1987], »Hierarchical Evolutive Systems: A mathematical model for Complex systems«. Bulletin of Mathematical Biology 49 (1987) 13; p.17.

[vii] MCLANE [1986], p.393.

[viii] EHRESMANN/VANBREMEERSCH [1986], p.19.

[ix] We adopt our example from MACLANE [1986], p.393.

[x] EHRESMANN/VANBREMEERSCH p.20.

[xi] In a sense there would be no geometrical location at all. The segments are in this restricted framework are only indexed disjoint point sets. Cf. EHRESMANN/VANBREMEERSCH p.20.

[xii] EHRESMANN/VANBREMEERSCH p.14seq.

[xiii] EHRESMANN/VANBREMEERSCH p.14. Cf p. 16, 18seq., p. 33.

[xiv] That is not quite correct, because one has to add the dimensions of the cycles of the homology group.

[xv] EHRESMANN/VANBREMEERSCH, p.33.

[xvi] EHRESMANN/VANBREMEERSCH, p.39seq.

On the Nature of Emergence
Interpretation and Understanding Revisited[1]

by Rainer E. Zimmermann (München/Kassel[2])

Abstract

The general ideas of transcendental materialism are outlined. It is shown how recent results of this theory can be interpreted in terms of a line of thought which is connected with the philosophies of Spinoza, Schelling and Bloch. In particular, results on the onto-epistemic mediation of the world (or worldly process rather) in the sense of Schelling are discussed in some detail. They are related then to a recent discussion about the ontological and epistemological state of the Combinatorial Hierarchy. It is found that this hierarchy can best be visualized, if it is interpreted in terms of modelling a concrete process of becoming. It is proposed that the introduction of an explicit background substratum is pursued in order to eventually achieve a pre-geometric foundation of such a process.

All mimsy were the borogoves,
And the mome raths Hograbe.

I

The general idea of a materialistic approach on the line *Spinoza - Schelling - Bloch*[3] (recently referred to as *transcendental materialism*) is the following: 1) Philosophy shows up as a *post hoc* orientation with respect to

---

[1] Contribution to ANPA 20, Cambridge, 1998.

[2] Lehrgebiet Philosophie, Fachbereich Allgemeinwissenschaften, Fachhochschule, Lothstr 34, D - 80335 München and Institut für Philosophie, Fachbereich Erziehungswissenschaft und Humanwissenschaften, Gesamthochschule-Universität, Nora-Platiel-Str.1, D - 34127 Kassel. e-mail: pd00108atmail.lrz-muenchen.de.

[3] In fact, this line of thought originates as early as in ancient Greek stoicism, and has dominated European philosophy a long time up to the "academy" of Charles the Great. It has been reformulated and continued within a modernized context by the Arab protagonists of what is generally referred to as the "Aristotelian Left". It has been transformed then by Bruno, Spinoza and others, finding its definite though still idealistic form in Schelling and leading forward through the materialistic turn of 19th century philosophy to the philosophy of Ernst Bloch. Cf. R.E.Zimmermann: The Utopian Function of Art and Literature in the Philosophy of Ernst Bloch. A Topic Revisited. Bloch-Almanach 15, 1996, 33-73. Also: The Ontological Perspective of Existential Dialectics. A Contextual Approach to Sartrean Thought. In: P.L.Eisenhardt, K.Palonen, L.Subra, R.E.Zimmermann (eds.), Modern Concepts of Existentialism. Jyväskylä Studies in Education, Psychology, and Social Research 102, 1993, 35-65.

the results science is offering for developing a general world view such that the *unification* of all partial fields of research dealing with various aspects of the world and the subsequent *foundation* of this unified view in terms of reconstructing the field of possibilities for such a world become motions which actually *follow up* the process of scientific research rather than laying the grounds for it (as has been the traditional view of the philosophical task). **2)** If so, the elementary view of the totality of the universe is being discussed in terms of cosmology as a global branch of theoretical physics. As on the ontological level, philosophy is still interested in the explicit difference between being and non-being, this means that philosophical results about this difference are *primarily based on models* being developed within theoretical physics. Hence, propositions made in the (axiomatic) foundation of these models (such as the cosmological principle) gain the quality of a "pragmatic ontologization" while being used on epistemic grounds in first place. In other words: The cosmological principle visualized as a proposition can be interpreted as a "synthetic proposition *a priori*", contrary to Kant's intuition. Consequently, part of what is called *non-being* (which is actually the field of possibilities for the becoming of a world such as ours) can be described in mathematical terms of the underlying physical theories (usually referred to as "pre-geometric theories" then). **3)** It can be shown that consequently, space-time-matter is practically the only attribute of (pre-geometric) substance. Hence, everything there is can be visualized as a product of the process which is defined as the unfolding (evolution) of the underlying potential. In particular, (human) thinking (in the sense of reflecting) is a product of this process, and therefore *thinking itself is a form of matter* (in the physical sense), though of a very complex kind. Reflecting about the world (e.g. doing research) can thus be visualized as worldly nature reflecting about itself: Nature is telling the results of the reflexion to itself, hence there is *a self-narrating aspect* in nature. This means that the world is *onto-epistemically mediated* in first place: The process in question is producing entities which are observing and interpreting the process itself. Hence, there are explicit boundary conditions as to what can be thought about (and *how* can be thought about it), because this is practically determined by the very nature of the process producing these thoughts among others. (Note here the explicit order of succession!) If therefore, a particular structure of the world is being revealed by means of actual research undertaken, then this can be visualized as an adequate mapping of things happening "truly", simply because the very activity of research is (though subsequent) part of the same process. (Provided research is being undertaken according to well-defined conventions which are of anthropological significance.) This is particularly true for the problem of (initial) emergence: Visualized as *a transition from*

non-being to being, the initial emergence of the universe is the realization of an approriate representation of substance while the subsequent (worldly) emergence of innovative structures is a consequence of the underlying process of evolution, and the interpreting and understanding of this emergent dynamism is also a part of the latter. Hence the onto-epistemic mediation mentioned above.
I will first (in section II) discuss shortly what the general problem of Schelling was when he introduced his "world formula". I show then (in section III) the relationship between this problem and the general problem of the (BNAK) combinatorial hierarchy (henceforth: CH). I conclude with a general discussion of how one could proceed (section IV).

II

The perceiving and cognitively grasping of the world is basically a problem of the adequate differentiation of being, that is: cognition straightforwardly points to the relationship between identity and difference. In particular, the problem of the identity of both identity and difference goes actually back to Hölderlin who introduces the concept of judgement (Urteil) in the double connotation of its meaning proper in the strict sense of predication, and of meaning "original division" (Ur-Teilung), at the same time.[4] In this concept, for Hölderlin, the concept of a relationship between subject and object is already included, and it implies a primordial unity (namely prior to this original division) of which both of them have been a part. In particular, in the famous proposition of Fichte, "I equals I" (Ich = Ich), for Fichte the first principle of his philosophy, Hölderlin notes an explicit threefold unity of a primordial identity which is broken up by the very beginning of reflexion (about oneself, in the sense of an elementary "cogito"): The I on the left-hand-side (the one which actually realizes itself as being) is an active *subject* while the I on the right-hand-side is a passive *object* (one which is reflected about). Hence, reflecting about oneself means to perceive oneself as subject and object at the same time (as difference that is) while knowing that both are identical in terms of the very notion of reflexion. The "I" which is reflecting is therefore an identity, but one which comprises both a difference and an identity of what is different: It is an identity of identity and difference. (In fact, this threefold division which renders the "identity equation" above into three different equations: - 1) I-subject ≠ I-object, 2) Unification (I-subject, I-object) = Identity, 3)

[4] F.Hölderlin: Urtheil und Seyn, in: Sämtliche Werke und Briefe, 2 vols., Wissenschaftliche Buchgesellschaft Darmstadt, 5th ed., 1989, I 840-841.

"I" *knows* of these facts, hence there is a unity of the first two equations establishing overall identity, - is a necessary condition of being able to reflect at all. Consequently, thinking of the world establishes an *ontological monism*, but an *epistemological dualism*!)

Schelling generalizes this idea by applying the basic principle to propositions of the form A = B, indicating that by an appropriate listing of propositions of this type knowledge about existential facts could be established. Indeed, this approach is very modern, in so far we today would think of a theory as being a set of propositions such that there is a derivational context governing the relationships among them. Starting such a listing with first principles of the Fichtean kind (with "Grundsätze") would establish then a systematic architectural structure modelling aspects of the real world. (It is in this very sense that the cosmological principle would turn up as such a first principle which gains the quality of a synthetic proposition *a priori*.) - Also, it would be obvious that the description of the world is started with a principle which is divided from the outset. Hence, true identity can only be established outside the world (in pre-worldly non-being that is). This also means that any systematization of the world is always incomplete due to its own self-reference, in the sense that at the world's origin (at the locus of initial emergence) there is a "blind spot" which cannot itself be part of a worldly theory. Schelling has tried to express this fact in terms of his "world formula", as I have shown at another place[5]. Hence, Schelling's "world formula" not only serves the purpose of shedding some more light on the problem of initial emergence, but also deals with preparing a semiological (or: semiotic) evaluation of the onto-epistemic foundation of the worldly and its mediation with the latter's evolution, respectively.[6] I will not repeat the explication of this world formula here, but it suffices to note that a prime aspect of this discussion is this fact of the irreducible singularity of self-reference. This means that the beginning of the world is hidden in the actual transition from non-being to being, and it is Hogrebe who locates *a cosmological meaning* in this: In the external unity he visualizes the singular which structures the dynamics of the self-organizing universe out of its own irreducebility and which ex-

[5] R.E.Zimmermann: Aesthetics as a Semiology of Nature. On the Unity of Schelling's Substance-Metaphysics. System & Struktur IV/2, 1996, 151-173.- See in more detail id.: Die Rekonstruktion von Raum, Zeit und Materie. Moderne Implikationen Schellingscher Naturphilosophie. Lang, Frankfurt a.M. etc., 1998, section 6.2, 193sqq.

[6] I use here the technical connotation of "worldly" in the sense of what "is belonging to the world", trying to translate the German "das Welthafte" which can also be called "das Weltliche". But I know that Ted Bastin does not like this usage of the word. However, in English there is apparently only one word appropriate for this which is "the worldly". Although this is often used in the sense of "mundane" or "profane", respectively (which in German is then"weltlich" exclusively). Roget's Thesaurus leaves open the possibility to use it in this technical sense of referring to the world as the Universe. (Penguin (new) edition, 1975 reprinting, 127, no. 321 in contrary to no. 319)

presses this fact in terms of symmetry breakings and phase transitions (being worldly traces of this hidden beginning). As Hogrebe states, the beginning is displaced (verdrungen) in this sense in everything what exists. Hence, there is always something ungraspable which looms into the world (from outside) and contaminates the obvious unity of the worldly. It is a "fundamental inconsistency" which "within all consistent situations and relationships is lurking foundation, and thus epistemic exaction."[7]

III

If now Clive Kilmister refers to a "becoming of entities" in the sense of Whitehead rather than to a "becoming *known* of entities" in the process of enlarging the region of the known world as it has been defined for explaining the CH in earlier work of himself and Ted Bastin[8], then it is necessary to examine Whitehead's standpoint first with a view to what we have explained above. The problem is possibly the position of Whitehead's theory with respect to the outlined transcendental materialism: Although Whitehead locates the motivation for his "complete cosmology" in the vicinity of speculative philosophy in the traditional sense, and although he stresses the relevance of the heuristical method and argues very much on the line of Spinoza[9], he nevertheless misses the central point of the relationship between non-being and being: He would like to categorically confront the concept of substance[10], and he establishes therefore his "actual entities" as "last real things from which the world is composed"[11]. However: If the real world is indeed a process, and this process is the becoming of actual entities such that the latter are also the real (actual) concrescence of many potentials being immanent in what he calls prehensions (as relational aspects among processes)[12], then he has to refer to a "field of possibilities" (a potential) which is with respect to what it is potential of *non-being*. In so far it is located outside the world (because it *is not*). And

---

[7] W.Hogrebe: Metaphysik und Mantik, Suhrkamp, Frankfurt a.M., 1992, § 19, 111-118.- Cf. id.: Prädikation und Genesis, Suhrkamp, Frankfurt a.M., 1989, 116. (Partially paraphrased here.) In fact, Hogrebe relates the necessity for art (doing the same as philosophy does, but with different means and pointing to different objectives) to this epistemic exaction.
[8] I mainly relate to Clive's Bielefeld paper: Physical Interpretation as Emergent, in: R.E.Zimmermann, A.Müller, K.Mainzer (eds.), Natura naturans, Topoi of Emergence, ZiF Bielefeld 1997, to be published. See also C.W.Kilmister: Discrimination with Aspect, ANPA 19 (1997).
[9] A.N.Whitehead: Process and Reality. An Essay in Cosmology. Macmillan, 1929. Here quoted according to the German edition (Suhrkamp, Frankfurt a.M., 1987) which is also aware of what is usually referred to as the "corrected edition" (N.Y., 1978). Note especially: 22-34, 38.
[10] Cf. H.Holzhey: Das Postulat eines neuen Naturbegriffs, in: id., A.Rust, R.Wiehl (eds.), Natur, Subjektivität, Gott, Suhrkamp, Frankfurt a.M.,1990, 18-40, here: 26.
[11] Process and Reality, op.cit., 57sq.
[12] Ibid., 64 (I & ii).

hence it is what is traditionally called substance.[13] So Whithead's approach is not really more abstract (or more general) than the metaphysics outlined above, in the sense that there would be a (new) integrated notion of process incorporating both the becoming and the knowing of this becoming, respectively. Instead, there are two possibilities: Either the concept of substance is the same (then a new theory is not necessary), or Whithead refers to worldly existence only, when talking of "actual entities." (But then he would have to explain their emergence out of seemingly nothing - in contrary to the philosophical "energy theorem": *Ex nihilo nihil fit.*) A similar argument is actually valid with respect to his proposal to get rid of the predicative structure of language.
Hence, returning to the CH, we remain either within the framework of substance metaphysics: The CH is some model then, which shall represent self-organizing processes. Then it is indeed *human made* (that is, it is produced by entities which are products of the world process themselves). Hence, in this case, the CH is mapping something which probably has already existed when there were no humans yet. If so, the process of producing the CH is an abstract model for another (concrete) process which should be defined in terms of space-time-matter. (Then we have returned to an old question I have introduced at the 1985 Whitsun Meeting at Cumberland Lodge. And it is unlikely to indicate what physical process this could be, given the fact that the model works basically in terms of a generation procedure for numbers only. So far as I know there has not yet been given a presentation which could explain the CH in terms of category theory or topoi - because this would actually introduce inherent process structures when writing down the mathematical formalism!)
Alternatively, when referring to Whithead e.g., there would be no satisfactory visualization of what the CH could actually be. (In other words: An entity does *not* really enter the theoretical analysis as a symbol which will

[13] In fact, there is another ambuigity in this approach, because it is not clear whether the universe is really in one piece (a totality) governed by general laws (being also a unity), cf. op.cit., 65 (v). The main problem is obviously one of the individualization of the affections of substance (its modi that is), a problem common in the theory of Leibniz which has been based on a critique of Spinoza's theory and on which Whitehead relies strongly. But as I have shown at another place, Leibniz misunderstood the ge-neral conception of Spinoza with respect to the question whether substance would be divisable or not. Cf. R.E.Zimmermann: Kontinuum und Diskontinuum. Von der spekulativen Logik Leibnizens zur spekulativen Physik. In: Leibniz und Europa (6th Int. Leibniz Congress Hannover), 1994, part I, 832-839. Hence, actual entities in the sense of Whitehead would be nothing but perspectively restricted representations of the attributes of substance. In so far, there is no innovative input which Whitehead could offer here. This may be the reason that his German translator (H.G.Holl) notes that "Whitehead's philosophy is basically Platonian." (op.cit., 643, with respect to Leibniz: 646) There are also theological connotations in Whitehead's approach which prevent the clear definition of substance. (Traditionally, God appears as an entity beyond worldly entities, hence takes the role of substance. In referring explicitly to God in terms of his basic propositions, Whitehead implicitly reintroduces the substance he would like to get rid of in first place.)

be its name, when it "becomes" in Whitehead's language, but it does enter this analysis *only, if* there is an observer (human or not) who can perceive it and reflect about it.)[14] It can be easily seen, I think, that a formalized process of observation can indeed constitute something like a "discrimination operator" D which when applied to two given entities A and B could reveal wether the one or the other has been there before or not. Possibly, this can also be achieved, if A and B are (complex) propositions. (In that case, D could help to decide e.g. whether these propositions are of analytical or synthetic type.) And it is also clear that it would be reasonable to expect some basic isomorphism between the mathematical structure expressed in terms of the procedure in which D is applied and acts on other given structures, on the one hand, and a (real) physical process which is mapped by this very procedure. This idea has been taken up again by Tegmark recently.[15] But it is far from clear how this could be visualized in the absence of observers. If there *are* observers however, then we are back to the picture of a modelling process.

IV

So how should we proceed? At an earlier occasion, I have related the CH to the elementary set of "anthropic invariants" which govern the processes of the universe in a fashion of fine tuning.[16] The idea was that if these invariants would be produced as the outcome of the generating procedure for the CH, then one could compare this generating with a model for the transition taking place between non-being and being at the locus of initial emergence. But then it would be necessary to specify the substratum which serves as an abstract "background" for this transition. In fact, a pregeometry would have to be specified in detail to show the explicit form of the transitional process. And this would make necessary far more structure than numbers can have at first sight. (Note that the original system of spin networks introduced by Penrose, which today is used in a variety of fields including strings and knots, dealt with (spin) numbers only, but started from an inherent though simplified particle picture establishing a kind of

---

[14] A third possibility would be to proceed in terms of numerology. As Tegmark has recently shown however, this is a route where success is highly unlikely, esepcially, if the analysis of heavy element production in stars were to reveal that the "archipelago of habitability" (in terms of constraints to world models) covered only a tiny fraction of the relevant parameter space. See his: Is "the theory of everything" merely the ultimate ensemble theory? Preprint April 1997, IAS, Princeton. (p. 28)
[15] See note 14.
[16] Cf. my ANPA 12 contribution (1990). See also my: The Anthropic Cosmological Principle: Philosophical Implications of Self-Reference. In: J.L.Casti, A.Karlquist (eds.), Beyond Belief (Abisko Summer Workshop 1989), CRC, Boca Raton, Ann Arbor, 1991, 14-54.

theoretical background for the exchange procedure of these spin numbers which eventually led to a superposition approximating Euklidean space concepts.)

What I would like to suggest is the following: If we assume that the set of anthropic invariants can be reproduced as an outcome of the full development of the CH[17], and that therefore the production procedure leading to the full CH describes a region in between the "ground state" (substance as non-being or potential, or field of possibilities for a future development) and the "initial state" of the universe (when it has emerged with the full set of significant invariants), then an appropriate physical process should be described which corresponds to what has been modelled so far. This also includes the description of the ground itself (of the foundation of the world, in terms of its mathematically formalizable part). This could be achieved when choosing as first orientations the available pre-geometric models such as strings, loops, knots, or twistors (including heaven theory).

The alternative would be to use the generation of the CH as foundation of these invariants in purely worldly terms (that is without leaving the world towards a pre-geometric state at all). But then it would be difficult to understand what this foundation could actually tell us, because the beginning of the procedure would not look like an assembly of necessary "first principles". (This was I believe Eddington's original idea: to derive physical constants as numbers from first principles. I cannot recognize a first principle so far.)

As a final remark I should note that according to my opinion, Tegmark's approach is a little too Platonian, in so far he turns the onto-epistemic principle over and creates an *epistemo-ontological* principle which appears to leave out the fact that reflecting entities such as humans (SAS in his wording) have been produced in the universe *after quite a while* and that physical processes should have been *very concrete* until then in order to make such a production possible at all. I prefer Trifonov's approach indicating that there is a close interrelationship between observers observing the world on the one hand, and what is being observed on the other, such that it is the explicit logic the observers choose for describing (modelling) their observations which determines what is actually observed. (The reason being that within the evolution of observers a choice of logics has been

[17] So far as I can see this is not a trivial problem: The four coupling parameters plus at least one mass ratio plus the cosmological parameter should be derivable. Although there is a large number of physical constants having been produced in terms of ANPA West by means of the program universe (P.Noyes et al.), I cannot decide whether all these relevant numbers can be actually produced from the outset, and if yes, on what level of the hierarchy they are showing up.

developed as part of the evolution of thinking which is - as we have seen before - nothing as a part of this very production process in first place.)[18]

1 Contribution to ANPA 20, Cambridge, 1998.
2 Lehrgebiet Philosophie, Fachbereich Allgemeinwissenschaften, Fachhochschule, Lothstr 34, D - 80335 München and Institut für Philosophie, Fachbereich Erziehungswissenschaft und Humanwissenschaften, Gesamthochschule-Universität, Nora-Platiel-Str.1, D - 34127 Kassel, e-mail: pd00108atmail.lrz-muenchen.de.
3 In fact, this line of thought originates as early as in ancient Greek stoicism, and has dominated European philosophy a long time up to the "academy" of Charles the Great. It has been reformulated and continued within a modernized context by the Arab protagonists of what is generally referred to as the "Aristotelian Left". It has been transformed then by Bruno, Spinoza and others, finding its definite though still idealistic form in Schelling and leading forward through the materialistic turn of 19th century philosophy to the philosophy of Ernst Bloch. Cf. R.E.Zimmermann: The Utopian Function of Art and Literature in the Philosophy of Ernst Bloch. A Topic Revisited. Bloch-Almanach 15, 1996, 33-73. Also: The Ontological Perspective of Existential Dialectics. A Contextual Approach to Sartrean Thought. In: P.L.Eisenhardt, K.Palonen, L.Subra, R.E.Zimmermann (eds.), Modern Concepts of Existentialism, Jyväskylä Studies in Education, Psychology, and Social Research 102, 1993, 35-65.
4 F.Hölderlin: Urtheil und Seyn, in: Sämtliche Werke und Briefe, 2 vols., Wissenschaftliche Buchgesellschaft Darmstadt, 5th ed., 1989, I 840-841.
5 R.E.Zimmermann: Aesthetics as a Semiology of Nature. On the Unity of Schelling's Substance-Metaphysics. System & Struktur IV/2, 1996, 151-173.- See in more detail id.: Die Rekonstruktion von Raum, Zeit und Materie. Moderne Implikationen Schellingscher Naturphilosophie. Lang, Frankfurt a.M. etc., 1998, section 6.2, 193sqq.
6 I use here the technical connotation of "worldly" in the sense of what "is belonging to the world", trying to translate the German "das Welthafte" which can also be called "das Weltliche". But I know that Ted Bastin does not like this usage of the word. However, in English there is apparently only one word appropriate for this which is "the worldly". Although this is often used in the sense of "mundane" or "profane", respectively (which in German is then"weltlich" exclusively), Roget's Thesaurus leaves open the possibility to use it in this technical sense of referring to the world as the Universe. (Penguin (new) edition, 1975 reprinting, 127, no. 321 in contrary to no. 319)
7 W.Hogrebe: Metaphysik und Mantik, Suhrkamp, Frankfurt a.M., 1992, § 19, 111-118.- Cf. id.: Prädikation und Genesis, Suhrkamp, Frankfurt a.M., 1989, 116. (Partially paraphrased here.) In fact, Hogrebe relates the necessity for art (doing the same as philosophy does, but with different means and pointing to different objectives) to this epistemic exaction.
8 I mainly relate to Clive's Bielefeld paper: Physical Interpretation as Emergent, in: R.E.Zimmermann, A.Müller, K.Mainzer (eds.), Natura naturans, Topoi of Emergence, ZiF Bielefeld 1997, to be published. See also C.W.Kilmister: Discrimination with Aspect, ANPA 19 (1997).

Page 1

[18] More details for a unified world view according to the above mentioned line of thought are given in the forthcoming: The Klymene Principle. A Unified Approach to Emergent Consciousness. (To be published in 1998.)

9 A.N.Whitehead: Process and Reality. An Essay in Cosmology. Macmillan, 1929. Here quoted accor-ding to the German edition (Suhrkamp, Frankfurt a.M., 1987) which is also aware of what is usually referred to as the "corrected edition" (N.Y., 1978). Note especially: 22-34, 38.
10 Cf. H.Holzhey: Das Postulat eines neuen Naturbegriffs, in: id., A.Rust, R.Wiehl (eds.), Natur, Subjektivität, Gott, Suhrkamp, Frankfurt a.M.,1990, 18-40, here: 26.
11 Process and Reality, op.cit., 57sq.
12 Ibid., 64 (I & ii).
13 In fact, there is another ambuigity in this approach, because it is not clear whether the universe is really in one piece (a totality) governed by general laws (being also a unity), cf. op.cit., 65 (v). The main problem is obviously one of the individualization of the affections of substance (its modi that is), a problem common in the theory of Leibniz which has been based on a critique of Spinoza's theory and on which Whitehead relies strongly. But as I have shown at another place, Leibniz misunderstood the ge-neral conception of Spinoza with respect to the question whether substance would be divisable or not. Cf. R.E.Zimmermann: Kontinuum und Diskontinuum. Von der spekulativen Logik Leibnizens zur spekulativen Physik. In: Leibniz und Europa (6th Int. Leibniz Congress Hannover), 1994, part I, 832-839. Hence, actual entities in the sense of Whitehead would be nothing but perspectively restricted representations of the attributes of substance. In so far, there is no innovative input which Whitehead could offer here. This may be the reason that his German translator (H.G.Holl) notes that "Whitehead's philosophy is basically Platonian." (op.cit., 643, with respect to Leibniz: 646) There are also theological connotations in Whitehead's approach which prevent the clear definition of substance. (Traditionally, God appears as an entity beyond worldly entities, hence takes the role of substance. In referring explicitly to God in terms of his basic propositions, Whitehead implicitly re-introduces the substance he would like to get rid of in first place.)
14 A third possibility would be to proceed in terms of numerology. As Tegmark has recently shown however, this is a route where success is highly unlikely, esepcially, if the analysis of heavy element production in stars were to reveal that the "archipelago of habitability" (in terms of constraints to world models) covered only a tiny fraction of the relevant parameter space. See his: Is "the theory of everything" merely the ultimate ensemble theory? Preprint April 1997, IAS, Princeton. (p. 28)
15 See note 14.
16 Cf. my ANPA 12 contribution (1990). See also my: The Anthropic Cosmological Principle: Philosophical Implications of Self-Reference. In: J.L.Casti, A.Karlquist (eds.), Beyond Belief (Abisko Summer Workshop 1989), CRC, Boca Raton, Ann Arbor, 1991, 14-54.
17 So far as I can see this is not a trivial problem: The four coupling parameters plus at least one mass ratio plus the cosmological parameter should be derivable. Although there is a large number of physical constants having been produced in terms of ANPA West by means of the program universe (P.Noyes et al.), I cannot decide whether all these relevant numbers can be actually produced from the outset, and if yes, on what level of the hierarchy they are showing up.
18 More details for a unified world view according to the above mentioned line of thought are given in the forthcoming: The Klymene Principle. A Unified Approach to Emergent Consciousness. (To be published in 1998.)

THE FOUNDATIONS OF SPECIAL RELATIVITY AND ITS CRITIQUE[1]

By IRVING STEIN

ABSTRACT: The introduction of a unique kind of binary random walk object allows us to define the concept of an object so that we can derive the space-time and energy-momentum Lorentz Transformations. From a critique of this concept we show that it is then possible to derive both the Schroedinger and Dirac equations.[2]

1 INTRODUCTION

In classical physics, either Newtonian or Einsteinian, space and time are realities independent of each other; i.e., there is no correlation or functional relation between the values of one and the other. Objects exist "in" space and time and are defined as space functions of time "associated" with a mass; that is, it is in the existence of objects that space and time find a correspondence. Space, time, mass, and object

[1] This work is an extension of the work presented in my book, The Concept of Object as the Foundation of Physics, Peter Lang, 1996

[2] This work is restricted to a single dimension, non-interacting object.

are taken as fundamental realities in terms of which other concepts such as velocity, etc. are defined.

The functions themselves are analytic, thus resulting in a deterministic physics. The cardinality of the sets of both space and time points or values is nondenumerable; however, the condition of analyticity allows the definition of an object to be specified by a denumerable infinity of conditions within an interval, $\Delta t > 0$; this can be done either by the specification of the derivatives at a given time or by specification of the function values.

However, from such an ontology it is not possible to derive the Theory of Special Relativity; in order to establish the Theory of Relativity it is required to introduce phenomena such as the constancy of the velocity of light in all galilean frames or the experimental result that no object apparently has a speed greater than that of light; there is nothing in the classical concepts of space, time, or object from which a maximum speed, the essential basis of relativity, can be inferred. That is, whether or not there exists radiation, or for that matter, gravitational waves, the Lorentz Transformation still holds. A space-time or energy-momentum Lorentz Transformation is of sufficient generality so that in pre-quantum physics it cannot depend on or be a consequence of any particular or contingent phenomena but must arise out of the very nature of an object itself in space and time. That is, it is not that the Theory of Relativity has its source in any particular phenomena, but rather that all phenomena that are defined in terms of space-time or energy-momentum must satisfy the Lorentz Transformation. In this sense, the source of the theory is yet to be found. This point is emphasized by Einstein himself in a letter to Sommerfeld[3]: "A physical theory can be satisfactory only if its structures are composed of elementary foundations. The theory of relativity is just as little ultimately satisfactory as, for example, classical thermodynamics was before Boltzmann had interpreted the entropy as probability." Einstein, as did Poincare, called such a theory of elementary foundations a "constructive" theory. . In this sense there still remains the problem of re-understanding the nature of these concepts and transforming them so that it is they themselves that give rise to the theory of relativity; this is the meaning of what Einstein meant by a constructive theory; and this is what we do in this work.

The amazing thing about such a theory as developed here is that not only can we then derive the theory of relativity, but find it as the necessary bridge to quantum mechanics.

Such classical concepts as space, time, and classical object as described above cannot give rise to either relativity or quantum mechanics. Relativity theory arises not out of classical physics but as a restraint, arising out of electromagnetism, on the classical concepts of space and time. However the existence of electromagnetism, as stated above, is irrelevant to the classical concepts of space and time. The postulate for the Theory of Special Relativity that all objects must have a speed less than that of electromagnetic waves such as light tells us nothing about the nature of space,

[3] Quoted in Jack Stachel, Einstein' Miraculous Year, Princeton University Press, 1998, p. 19

time, or classical object from which such a limitation must arise

In this work I investigate what such a revision must be so that from it can arise the space-time and momentum-energy Lorentz transformations. It will be shown that this revision is based upon the concept of what I call precursor objects. From an analysis based on the nature of these objects we will be able to derive the Lorentz Transformation, both the space-time and, after showing the necessity of the concept of mass, the momentum-energy forms.

The contradiction between classical and quantum theory is even more glaring. There appears to be no way that quantum mechanics can arise from the classical concepts of space, time, and object. In quantum mechanics classical concepts , such as frames of reference, position, and velocity have no meaning except insofar as wave functions are referred to classical kinematic variables. Thus, although the wave functions of quantum mechanics can be defined in terms of position and often refer to position and galilean frames of reference, the objects that quantum mechanics deals with do not normally have the properties of position or velocity, and therefore can not be referred to frames of reference, either spatial or velocidal; in fact, the very concept of object needs to be revised, not only for quantum mechanics, but also for relativity.

Thus, despite the overwhelming success of the theory for the derivation of relativity, it is, at best, only a basis from which to derive quantum mechanics. Therefore, we subject it to a critical analysis, removing all assumptions, assumptions that also exist in classical, or pre-quantum physics, that cannot be justified. The resulting bare-boned physics, based on what I call nonspace, turns out to be the ontology, the basis, of quantum mechanics. That is, the precursor objects turn out to be the precursor of the even more fundamental concept of nonspace, which is nothing but all possible Feynman integral "paths," and thus the spin of objects. These "paths" derive from the precursor objects. From this ontology and the nature of measurement we indicate how, first the Schroedinger, and then the Dirac, equations were derived in reference [1].

II RELATIVITY

In order to develop the theory of relativity arising directly out of a transformed concept of object we must first clearly understand the fundamental nature of the Theory of Relativity. It is not so much a new "relativity principle," since it also holds in Newtonian physics, it is rather based on the fact that all objects have a maximum speed; such a fact derives either from experiment or from the logical consequences of electromagnetic theory as Einstein showed. Based on this restriction it is then found that the principle of relativity holds for all classical situations, even for quantum mechanical wave functions although not for the quantum mechanical objects themselves, as stated in the introduction. It is this apparently very strange restriction that is put onto the concept of object from which the Theory of Relativity arises. The question for us here is: is it possible to modify the concept of object so that the very concept of object itself will give rise to the Theory of Relativity? The answer is yes. We do this by introducing the concept of a precursor object. But the precursor object not only allows us to derive the Theory of Relativity, but it also is fundamental to the derivation of quantum mechanics.

1. We define a precursor object as a binary object, an object having two values, +c or -c.

2. A value of the precursor object is defined at each value of time. Thus, at each value of time there is defined a value + c or - c value of the precursor object.

3. These values are distributed randomly over the set of time points so that the mean over any time interval is zero and the standard deviation over any interval is c.

4. I now define on any time interval, $\Delta t > 0$, a sequence of precursor values containing a fraction, p, of + c precursor values and a fraction, q, of - c precursor values, so that p + q =1. The average, v, of the sum of these values, is then defined as $v(t_0, \Delta t) = p(t_0, \Delta t) - q(t_0, \Delta t)$, where t_0 is a limit point.

5. Furthermore, we define the function, $p(t_0, \Delta t)$, so that it has a limit at $t = t_0$; that is, $p(t_0) = \lim_{\Delta t \to 0} p(t_0, \Delta t)$ and

$v(t_0) = \lim_{\Delta t \to 0} v(t_0, \Delta t)$.

6. There are, of course, an infinity of sequences, $\{f(t_i)\}$, that

satisfy the condition $v(t_0) = \lim_{\Delta t \to 0} \left\{ v(t_0, \Delta t) = \lim_{n \to \infty} c \frac{\sum_i^n f(t_i, \Delta t)}{n} \right\}$

7. Now suppose that we wish to add two such sequences, $p_1(t_0, \Delta t)$ and $p_2(t_0, \Delta t)$. How is this to be done?

8. Just as in the usual formulation of classical physics where the sum of velocities is defined as the sum of two distances over the <u>same time</u> interval as it approaches zero. So similarly here the addition of any terms of the two sequences, $\{f(t_i)\}$ and $\{g(t_j)\}$ must also be at the same time; that is, i = j. Therefore, in order to add two sequences of distributions $p_1(t_0, \Delta t)$ and $p_2(t_0, \Delta t)$, we must chose the sequences so that they have like terms defined at the same times, $i = j$.

9. Since the sum of the two sequences is also to be a binary sequence it is required to define the addition of like terms such that $+c \oplus +c = +c$ and $-c \oplus -c = -c$. And since at any time instant the precursor value is **either + c or - c, there cannot be a + c from one sequence and a - c from the other sequence at any given time. Thus, in the sum, there is, once again, only + c and - c terms.** That is, in order to be able to add two sequences it is necessary to define a random binary object on the time instances as we have done.

10. We can now express p(t) and q(t), in terms of the average, v(t), where p(t) + q(t) = 1 for all t of the sequence;

$$p(t) = \frac{1}{2}\left[1+\frac{v(t)}{c}\right], \qquad q(t) = \frac{1}{2}\left[1-\frac{v(t)}{c}\right].$$

11. Now, as stated above, we define an element of the third sequence to be +c if there is a +c at the same time in both the first and second sequences. Thus the probability of there being a +c from each of the first two (independent) sequences is p_1p_2 and the probability of there being a -c is q_1q_2. Since an element in the third sequence is +c if there is a +c in each of the first two sequences and –c if there is a –c in each

of the first two sequences, and no other contributions to the sum sequence, we conclude that the fractions of +c's and –c's in the third sequence are:

$$p_3 = \frac{p_1p_2}{p_1p_2+q_1q_2} \text{ and } q_3 = \frac{q_1q_2}{p_1p_2+q_1q_2}.$$

12. This, of course holds not only within the interval, Δt, but also at the limit point, t_0. Thus, at some point of time, t_0, there exists not only a precursor value, +c or –c, but also the limits of the average of these values, one each for any sequence, and also of the sum. The limit of the average of the sum at the limit point, t_0, can now be evaluated from the probabilities p_3 and q_3. In expressing the above equations in terms of the averages we get, $v_3 = \dfrac{v_1+v_2}{1+\frac{v_1v_2}{c^2}}$.

13. This equation is reminiscent of the "sum of addition formula" of special relativity. And it is true that the precursor object in being able to produce this formula has penetrated to

the heart of the physics of special relativity. However, a Lorentz transformation is first a space-time transformation and we have no concept of space yet. Therefore we require a way to develop the concept of space. We do this by identifying the concept of the average of the precursor values, $v(t_0)$, with the concept of the classical velocity, $v(t_0)$; (we now use the same notation for both) However, classical physics requires the velocity to be continuous (in fact, analytic). Thus, our concept of the average of the precursor values, $v(t_0)$, must also be continuous (analytic). This means that there must be an interval, Δt, in which the average--- or that which is now the same--- the velocity, is defined and continuous. Thus, its integral exists and we can write, $\Delta x = \int v(t)dt = x - x_0$. Since the integrand is the velocity, the integral then is a space interval; we see that position is relative.

14. The limit of the average of a sequence of precursor objects, in now having the restriction on it as an analytic space function of time, thus defines not only velocity but a space interval and thus can define a classical relativistic object except for one further required development -- the v's developed above are not relative to other v's; they are absolute In order to define classical physics and derive the Lorentz Transformation, the concepts of galilean frames of reference and relative velocity must be introduced. We define the concept of the relative velocity, $v(t)$, by defining the value of one average function, $v_i(t)$, with respect to the value of another average function, $v_j(t)$. When defined, such a function will be designated $v_{i/j}(t)$.

15. The first step in such a process is to define the value of v relative to that of $v_0 = 0$ as having the same value; that is, $v_{1/0} = v_1 - v_0 = v_1$. Then if $v_1 = v_{1/0}$, $v_2 = v_{2/1}$, and $v_3 = v_{2/0}$, from the above formula for the absolute velocities we have, $v_{2/0} = \dfrac{v_{1/0} + v_{2/1}}{1 + \dfrac{v_{1/0} v_{2/1}}{c^2}}$.

In words, this says that the sum of two such functions, one, v_1, relative to $v_0 = 0$ and the other, v_2, relative to the first, v_1, gives a value $v_{2/0}$, by the above formula, of the second, v_2, relative to $v_0 = 0$.

16. If we now solve for $v_{2/1}$, we then get the relative value between <u>any</u> two functions in terms of their values relative to $v_0 = 0$, That is, $v_{2/1} = \dfrac{v_{2/0} - v_{1/0}}{1 - \dfrac{v_{2/0} v_{1/0}}{c^2}}$.

17. Therefore, we now can write, $v_{2/0} = \dfrac{v_{1/0} + v_{2/1}}{1 + \dfrac{v_{1/0} v_{2/1}}{c^2}}$.

18. But since the <u>absolute</u> average, v(t), is integrable, so is the <u>relative</u> average, $v_{i/j}(t)$. This allows us, as before, to define a variable, x, where $\Delta x = \int v_{i/j}(t)\,dt$. Thus, whether or not velocity is absolute or relative, position is relative. Then, if v(t) is constant, we can write, after integrating, $x - x_0 = v(t - t_0)$

19. Since, the concept of space has now been defined, we use it in its differential form to rewrite the above formula for relative velocity as, $v_{2/0} = \frac{dx'}{dt'} = \frac{v_{1/0} + \frac{dx}{dt}}{1 + \frac{v_{1/0}}{c^2}\frac{dx}{dt}}$, where $v_{1/0}$ is taken as a constant. We then write, $\frac{dx'}{dt'} = \frac{v_{1/0}dt + dx}{dt + \frac{v_{1/0}dx}{c^2}}$, and therefore, $dx' = \gamma[v_{1/0}dt + dx]$ and $dt' = \gamma[dt + \frac{v_{1/0}}{c^2}dx]$, where γ may be a function not only of $v_{1/0}$ but possibly also of x and t. We can show, however, that it is not a function of x or t in the following manner.

20. Recall from the meaning and definition of "relative" that $v_{i/j} = -v_{j/i}$. Therefore, we can write,

$v_{2/0} = \frac{-v_{0/1} + v_{2/1}}{1 - \frac{v_{0/1}v_{2/1}}{c^2}}$. Solving for $v_{2/1}$, we get

$v_{2/1} = \frac{dx' - v_{1/0}dt'}{dt' - \frac{v_{1/0}}{c^2}dx'} = \frac{dx}{dt}$, from which we get

$dx = \gamma\ [dx' - v_{1/0}\,dt']$ and $dt = \gamma\ [dt' - \frac{v_{1/0}}{c^2}dx']$.

Substituting these expressions for dx and dt into the equations, $dx' = \gamma[v_{1/0}dt + dx]$ and $dt' = \gamma[dt + \frac{v_{1/0}}{c^2}dx]$, we get $\gamma = \frac{1}{[1 - (\frac{v_{1/0}}{c})^2]^{1/2}}$.

If $v_{1/0} \equiv v$, then integration gives us, $x' - x_0' = \gamma[(x - x_0) - v(t - t_0)]$, $t' - t_0' = \gamma[(t - t_0) - \frac{v}{c^2}(x - x_0)]$

If we take $x_0', t_0' = (0, 0)$ then we simplify to $x' = \gamma\ [x - vt]$ and $t' = \gamma\ [t - \frac{v}{c^2}x]$.

Thus, based on the concept of a binary variable, the precursor object, I have been able to derive the space-time Lorentz transformation.

21. I now recapitulate what has been done so far:

On each element of the set of real numbers (time), $\{t\}$, is defined a random binary variable, $\pm$ c, which I call a precursor object. On each element of the set of real numbers, $\{t\}$, is also defined $v(t_0) = \lim v(t_0; \Delta t)$ where $\Delta t \rightarrow 0$. That is, $v(t_0)$ is the limit of the average of a denumerable set of $\pm$ c points in Δt as $\Delta t \rightarrow 0$. It is seen that $v(t_0) < c$ except where, with a probability of zero, $v(t_0) = c$. Thus, for any precursor variable sequence, that is, for any random binary "walk," we define a limit to the average of its values. The limit of the magnitude of such an average will always be less than c If $v(t_0)$ is now taken to be continuous it then can be identified as a velocity.

22. A function, p(t), is defined as $p(t) = \frac{1}{2}[1 + \frac{v(t)}{c^2}]$; this is the fraction of $v(t) = +c$ at points, $t \in \Delta t$, as $\Delta t \rightarrow 0$. If the function q is defined as $q = 1 - p$, so that $p + q = 1$, then q is the fraction, similarly, of $-c$ values at points, t, in Δt as Δt approaches zero; thus both v(t) and a precursor value are defined at every value of the set of reals, $\{t\}$.

23. However, it is still necessary to define the addition of two velocities. Suppose that we have two such functions defined over the same domain of real numbers, $v_1(t)$ and $v_2(t)$. How

can their addition be defined? I have shown that apparently the only reasonable way results in the formula,

$$v_3 = v_1 \oplus v_2 = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}}$$

24. The reason why $v_3 \leq v_1 + v_2$ clearly comes out in the definition of the addition of the c's. Although the precursor object, even if it could be defined, would be highly discontinuous, $v(t)$ can be continuous and even analytic. We assume that it is. In defining a function $v(t)$, or equivalently p (t), on the precursor object, we take a step towards defining the classical relativistic object. If we now define $v(t)$ or $p(t)$ as analytic, we then have taken the next step in defining the classical relativistic object.

25. The last step in defining the classical relativistic object is to show how to redefine its velocity function as a <u>relative</u> average function, $v_{i/j}$. The above addition formula is then shown to be able to be written in terms of the relative average function as

$$v_{2/0} = \frac{v_{1/0} + v_{2/1}}{1 + \frac{v_{1/0} v_{2/1}}{c^2}}, \quad \text{where } v_2 = v_{2/1}.$$

26. From this, since the $v(t)$'s are integrable, one can then define $dx = v(t)dt$ and show that that $x' = \gamma(x - vt)$, $t' = \gamma(t - \frac{v(t)}{c^2}x)$, where $\gamma = \frac{1}{[1-(\frac{v}{c})^2]^{1/2}}$, and v is one of the above velocities, taken as a constant.

27. It is in this manner that I put sufficient restrictions on the precursor object, a binary random walk object, to transform the time sequence of a precursor object into a classical relativistic object in classical Einsteinian space-time. Loosely, it can be said that the theory of special relativity is a consequence of the magnitudes of space and time intervals in the small being identical.

28. A classical relativistic object is defined as a mass "associated" with a space function of time, where space is derived eventually from an analytic function of time, $v(t)$, which is defined in terms of random binary variable, $\pm$ c. It is this basis from which the space-time transformation, $x' = \gamma(x - vt)$, $t' = \gamma(t - \frac{v(t)}{c^2}x)$, which we now call the Lorentz Transformation, is derived. No reference need be made to any concept other than that of the concept of the precursor object and the usual continuity restrictions on it. The concept of a classical Newtonian object is also defined as a mass "associated" with an analytic function of time, which, however, is not based on a precursor function. Therefore, in this case there is no limiting velocity and the space-time transformation is galilean.

29. Only when we get to quantum mechanics does the concept of mass arise from the very nature of the object itself. However, even before we get to quantum mechanics we can show that the Lorentz Transformation, when applied to more than a single object, requires the concept of mass. This requirement depends not directly on the precursor variable but on the Lorentz Transformation.

30. Suppose that we write the Lorentz Transformation of space-time points as a rotation in a Minkowski (pseudo-Euclidean) space. Then, the space-time four-vector, **u**, can be written as,

$$\sqrt{-1}\,\mathbf{u} = (\mathbf{r}, \sqrt{-1}\,ct),$$

and the normalized difference between any two space -time points as,

$$\sqrt{-1}\,\frac{\Delta \mathbf{u}}{\Delta u} = \sqrt{-1}\,\mu = (\gamma\frac{\mathbf{v}}{c}, \sqrt{-1}\gamma)$$

where μ is the unit four-vector.

Suppose that we have a system of two, μ_0, of two objects, μ_1 and μ_2 where $\mathbf{v}_1 = -\mathbf{v}_2$. Then if the mass is not introduced, we have,

$$|\mu_0| = \tfrac{1}{2}|\mu_1 + \mu_2| = \gamma.$$

This, of course, cannot be since the left-hand side of the equation is a constant and the right-hand side is a variable depending upon the velocity, v. It is at this point that another variable must be introduced. Therefore, instead of equation (9) we introduce the concept of mass and write,

$$m_0\mu_0 = \tfrac{1}{2}[m_1\mu_1 + m_2\mu_2].$$

Then, once again, if $\mathbf{v}_1 = -\mathbf{v}_2$ and $m_1 = m_2$, we have $m_0 = \gamma\, m$. Thus, if the system, as defined by m_0, is constant, then m, the mass of each object must be a variable. It is in this manner that the concept of mass is first introduced.

31. In summary: the concept of the highly discontinuous two-valued object, ±c, which does not occur in ordinary classical physics is the basis of classical relativistic physics. What has been done in this work so far is to find the "underpinning" to the classical concepts of space and (relative) velocity. Fundamental to and "underneath" the continuous nature of space and velocity is a highly discontinuous discreteness. This discreteness arises from the random binary precursor object.. Because of this, the construction of the concept of object as done here results in the Lorentz transformation without any reference to any particular phenomena other than the very "nature" of itself. However, as will see in the next section, the precursor object is just the classical incursion to an even more fundamental reality, that of nonspace.

III QUANTUM MECHANICS

32. Nevertheless, there are serious problems in the theory presented above insofar as it still is a deterministic theory, even with the "underpinning" precursor variable. A deterministic theory, by its very nature, is dependent upon making a statement about the present based on claims on the future[1] even though it appears to be the other way around. Thus, in the theory presented above, all of the v(t) averages approach a limit, $v(t_0)$, at times, t, not only before, but also after t_0. How can something in the future determine something preceding it? Likewise, if $v(t_0)$ is to be continuous or analytic at $t = t_0$, then there are restrictions on $v(t_0)$ not only from a former but also from a later time. Furthermore, if

v(t) is analytic in some region, then the value of the function at any point in the region determines the function not only in the future but also in the past. Such a physics has little justification outside of the fact that apparently it has worked very well, at least until the discovery and development of quantum mechanics. Another objection to the above theory is that for a given relativistic object, even for one whose entire trajectory is defined over all of time, or even only at an instant of time, there are an infinity of precursor object sequences from which the same $v(t_0)$ could arise; the theory at best, then, is incomplete—there is no way in this theory that a unique velocity (sequence) function can be determined or is there any rationale why it should be one function instead of another.

33. A third objection is that although the theory requires objects to have mass, it requires the existence of two or more objects to show this; that is, the existence of mass is not a consequence of the nature of the object itself.

34. If therefore one gives up all arbitrary preferences for one result instead of another, if one then allows all possibilities, that is, if one attempts to be as consistent as possible in constructing a theory that does not make the present depend on the future and is as complete as possible, then what kind of theory would result? Would there even be the possibility of a theory? And how would it be done?

35. In order to produce a theory with no preferences, that is, a theory where the present is not dependent on the future or the past and is furthermore complete, there could be no

sequence, no function, $v(t_0)$, at the basis of the theory; <u>any</u> function, even one with p = q = 50% has a preference; this preference being simply that 50% of the ± c values are preferred +c and 50% are preferred -c. No preference means that at <u>each</u> value of time, the object is <u>both +c and -c. That is, at every value of time, the object assumes all possible values.</u> No average limit or, for that matter, no limit of any kind would be approached. Thus, since there would be no average or instantaneous velocity, there could be no space. An object in such a state is called a <u>nonspace</u>.

Eventually, this must be the ontology, or "non-ontology" upon which physics is based. But since the only kind of measuring instruments available are classical objects, objects that make position coincidences or distance measurements, any definition or expression of the nature of the nonspace must be in terms of classical objects.

36. This now is the clue to the relationship between the precursor variable and the concept of object as nonspace. The precursor object is only the precursor to an even more basic concept; that is, that at every value of time, at every instant of time, there is a not a precursor object of a value +c or -c. but an object, which I call a nonspace, defined by the fact that at every instant of <u>imaginary</u> time its values are both +c <u>and</u> -c. On the other hand the <u>precursor</u> object value, +c or -c, is one of the nonspace or object values when a successful measurement is made on it. If no measurement is made then each nonspace element of +c and -c mitoses into a pair of +c and -c elements continuously and randomly in <u>imaginary</u> time so that the number of nonspace values at any instant of

imaginary time increases linearly. Only if a successful measurement is made do space and time reappear and the process begins anew. Thus, the precursor object values only appear as a result of measurement. This is how the precursor object leads us to the fundamental reality of the nonspace.[1] It should be noted that the "width" of the wave function solutions of the Schroedinger equation with no potential, even though a non-relativistic equation, increases linearly with time, thus indicating, nevertheless, a relativistic source.[4]

37. A successful measurement is defined as the position coincidence of two objects; an unsuccessful measurement is defined as a failed position coincidence of two objects where it was possible to have a successful measurement. If there is a successful measurement, the object that was a nonspace is no longer the <u>same</u> nonspace since it now does have a position in coincidence with that of the classical object.; that is, a classical object is a highly specialized nonspace. This means that now there is a distance defined, a distance between the original position and the final position. This distance divided by the elapsed time may, if we wish, be called the velocity. But no velocity or position of the nonspace object existed during the time between the two position measurements. However, an unsuccessful measurement does not mean that nothing has occurred to the nonspace; on the contrary, the nonspace is transformed into another nonspace, one, however, that is not a space point. An example of this is an electron or other fundamental particle passing through two or

[4] See any first year text on quantum mechanics, such as Schiff's, Bohm's, et. al; also the derivation in reference [1].

more slits; the electron nonspace has been transformed to another nonspace other than a space point. Normally when I use the term nonspace, I will not mean a classical or space object.

38. The value of mass is defined by the average time of reversal of the nonspace for zero distance "moved" during any time. If this time of reversal is close to zero, the nonspace is essentially a classical object. Thus, between reversals of the nonspace, space and time are identical in magnitude, and the mass is defined by this space or time magnitude. That is, to this degree and in this manner, space, time, and mass are identical; they differ only in the way described.

39. In this way, by successful measurements, nonspaces become classical objects; if there is interaction during measurement so that the nonspace becomes part of the measuring agent; e.g. an electron becoming part of a molecular structure, it remains a classical object.

40. On the basis of the concept of nonspace defined here, both the Schroedinger and Dirac equations have been derived. The Schroedinger equation is derived on the basis that two successive position measurements made on an object define a single "step" in space; from this it immediately follows, as shown in [1], that the nonspace width or "dispersion" in terms of space increases linearly with time and not as the square root of time as in classical physics. From this, we then show that nonspace time is imaginary; the Heisenberg Uncertainty Principle then follows

as does the Schroedinger equation [2]. Using these results and a detailed counting of the continuous steps of the nonspace. one can then derive the Dirac Equation. As shown in [1], this work converges with Feynman's path integral concept and uses the significant calculations made by Jacobson and Schulman

41. However, it should be pointed out that all these equations do is, on the basis that the future can tell us absolutely nothing about the present (at least for noninteracting objects), translate the initial conditions into the results of future measurements. That is, the initial state of a nonspace object can be generally specified by an array of classical objects such as a grating. Then, if there is equal preference for all consequent states, which means that the future cannot determine the present, the probability density, and only the probability density, of the results of (position) measurements at any future time is determined. That is, the calculation of this probability density of position, which is expressed by the Schroedinger and Dirac equations, is not a statement about physics but rather the logical consequences of the physics stated above as a function of time. That is, from the lack of any preference and the consequent establishment of imaginary time and nonspace as the basic ontology of physics, one can derive the "kernel", the Green's type function that carries the initial condition of an object into the results of later measurements on the object. It is only the restriction on the average time of reversal of the nonspace, which defines the mass, which enters the derivation as physics.

42. The ± c values of the precursor variable are the eigenvalues of the relativistic velocity operator in quantum mechanics. Furthermore, in 3-dimensions the precursor objects are the basis of the concept of spin.[5]

43. The physics that results from the answer to all of these questions promulgated at the beginning of this section is quantum mechanics.

1V CONCLUSION

The source of both relativity and quantum mechanics has been discovered. This source is what I call nonspace, indicated by the concept of the precursor object, and is simply an object of "all possibilities" defined in terms of positions of classical objects and is inferred from the concept of the precursor object.. Since the reversal time of the nonspace, which defines the mass of an object, can approach zero, classical objects and space can exist and thus nonspace can be defined in the only way possible, that is, in terms of classical objects.

The proofs of the statements on quantum mechanics and detailed material on nonspace and measurement are presented in the book, The Concept of Object as the Foundation of Physics, 1996, Peter Lang. An excellent paper containing references to related work is V. A. Karmanov, Physics Letters A 174 (1993) 371-376 .

[5] T. Jacobson, J. Phys. A: Math. Gen. 17 (1984) 2433-2451

WIDER PERSPECTIVES - NATURE, COGNITION AND QUANTUM PHYSICS

Peter Marcer
53 Old Vicarage Green
Keynsham, BS31 2DH, UK.

1. INTRODUCTION

The following facts, emergent from the new interdisciplinery area of quantum information processing [1], preview a major paradigm change in science :-

a) in computer science, a physical theory - the quantum theory of computation - is now to be regarded as the theory of computation, replacing the mathematical/Turing theory as the correct one,

b) the nature of information is radically extended by the concept of quantum information, beyond what, until now, has generally been accepted in science to be the case,

c) information therefore becomes a new concept on a par with the accepted concept of energy, needing incorporation in understanding physics, and

d) as already experimentally validated, this incorporation radically changes the scientific understanding of how chemistry may be be performed - specifying new designs for chemical systems employing optimally controlled quantum signal induced, rather than approximately thermodynamically induced, chemical reactions [2,3,4].

The new paradigm, implies a wholly quantum world, where both classical models of physics and computation are at best approximations to physical reality, and where, for example, the recent experimental validation of quantum teleportation[5] is a striking affirmation of the holistic nature of physical reality revealed by quantum mechanics[6].

Thus, current understanding of all derived disciplines where the concept of information processing plays a leading role - such as molecular biology, biology, neurophysiology, cognitive science, medicine, etc, warrant re-assessment and maybe radical revision in the light of this paradigm change. **For example, a valid hypothesis is that the 1952-3 experiments of Miller and Urey into the origins of life (where within a closed flask, electric discharges through a 'primordial Earth atmosphere' for many days produced an amino acid containing chemical soup) were essentially correct, except that the possibility of optimally controlled quantum signal induced chemistry [3], was then unknown and is at the heart of life's evolution.**

2. THE PROGRAMME

From a programme of investigations already in progress, which concerns DNA, the prokaryote cell, the brain, and its essential sub-unit - the neuron -providing a new model of cognition, and the central challenge proposed, it becomes clear that no area of science may be immune from this extended understanding of information processing that quantum physics provides, and that the solution to many of the major unsolved problems in science, could now be within reach.

In the new paradigm, the only valid proofs (even for mathematicians!) are "engineered solutions" ie physical apparatus, be they manmade or biologically evolved, which demonstrate the process in question, as theoretically described. Thus brains as "chemically based computer design", have some information processing operations that are astounding from the engineering, technical viewpoint; physical operations which serve as an existence proof that carbon-based analogue chemical computation does work.

The investigations concern quantum holography (summarised below), for which functional magnetic resonance imaging furnishes an engineered solution [7]. They provide existence proofs ; ie descriptions of why the morphology and dynamics of the following chemically based computer designs, ie "engineered solutions", are the way they are: -

i) DNA - *describing how its morphology and dynamics is able to incrementally encode the three dimensional morphology and dynamics of the embryo of its organism*[8],

ii) the prokaryote cell - *describing how such cells function in an environmental niche, for the purposes of survival, as self organising, self replicating, self adaptive units, in accordance with an evolving record within the cell of the cell's history*[9],

iii) the brain as a cognitive system - *describing how it works as a fully distributed, massively parallel, synchronously partitioned processor, filter, memory, and multiple input/output system, which optimizes itself by self adaption/learning*[10], and

iv) the neuron of such a brain - *describing how it contributes to the brain's capabilities and massively parallel operation* [11]*and why this is quantum parallelism, such that, in relation to simultaneous activity in geometrically separate brain segments and to affecting of a change of functionality (ie the shutting down of activity in one set of segments and replacing this by that in another set), these can be done instantaneously, rather than requiring the use of co-ordinating signals from a central control or switching centre as in a classical machine.*

Biological evolved engineered solutions demonstrate, for example, that vision, visual images; sound, acoustic images, touch, tactile images, etc -senses- concern chemical simulation of physical reality, where (if quantum holography correctly describes them) adaption/learning, filtering, communication, input/output, memory etc take place directly by means of actual physical signals in the form of holograms; for example, immediate data capture is of a whole three dimensional vista/ 'movie' frame. No form of logical or mathematical representation is employed, simply a physical morphology and dynamics, which quantum holography prescribes, so that processing is top down or bottom up. This allows optimally efficient cognition by filtering of a whole, into its parts or perception by the assembly of a whole, from its parts, etc.

Other engineered or claimed engineered solutions, described by quantum holography are Mach-Zehner interferometry[12], *NMR quantum computation Jones, Mosca,and Hansen [25] and teleportation as performed by Zeilinger et al. and di Martini et al. [5].*

3. SCIENTIFIC BACKGROUND

Quantum holography describes the mechanisms of holography via the standard quantum mechanical formalism[12] specifying an emitter/absorber model equivalent to an incremental Feynman sum of histories approach[13]. It incorporates signal theory for generic three dimensional image processing and communication into quantum physics in acccordance with the symmetries of the nilpotent Heisenberg Lie group structure, algebra and nilmanifold[14]. These symmetries avoid the decoherence that would, in other circumstances, prevent the quantum interference, self-interference/entanglement necessary for the quantum holography. It does this by employing phase conjugation/ time reversal invariance[15], which are preconditions for self adaption, insensitivity to initial conditions, and what, in other contexts, is known as "stochastic resonance", such that inhomogeneity/ multiple scattering act, counterintuitively, as focussing mechanisms.

4. THE CENTRAL CHALLENGE

Biological systems are now simply "engineered solutions" for which DNA - itself an "engineered solution" - provides descriptions. Central to these descriptions to quote de Duve, a Nobel Laureate for cell biology[16], ".... the function of base-pairing is not simply a structural one. Its main role is communication. Amazingly these elementary relationships, the chemical base pairings (where A,U,T,G and C are adenine, uracil, thymine, guanine and cytosine respectively)

$$A = U \text{ or } A = T \qquad \text{and} \qquad G \equiv C$$

govern, through the two relatively fragile structures they embody, the whole of information transfer throughout the biosphere. They are truly the cipher of life."

That is, DNA base-pairing, the basis of the genetic code provides an "engineered solution" or basic machinery of communication fundamental to life on Earth. Thus, **the hypothesis that biological systems are quantum information processing based chemical computer designs, is reduced to experimentally validating that the machinery of base-pairing corresponds to four operations, a quaternary code,for the maintenance of entanglement/unitarity, namely:-,**

a)the unitary operation that simply does nothing or

b)either one of the three unitary transformations transfering one Bell state to another.

These unitary operations, the identity operation, the phase flip, the bit flip and the phase/bit flip, which are central to quantum error correction, and fault tolerence, describe for example an ability -quantum dense coding- to send twice as much information as can be sent classically with a two-state particle/system[1]. This, it is argued, must result in inevitable evolutionary advantage. The implication is that information with respect to both amplitude(bits) and phase is being transmitted as the basis of biological communication. A case, exactly analogous to the condition of 'quadrature' in holography[17], which allows full wavefront reconstruction, restoring the original 3 dimensional image carrying wave in both amplitude and phase; the same condition, which in quantum holography is guaranteed by the symmetries of the three dimensional representation G of the Heisenberg group, and where these symmetries imply avoidance of decoherence.

That is to say, quantum holography extends the classical model of holography for full wavefront reconstruction (upto a particular resolution) with complex amplitudes into the quantum domain. Thus the holographic information needed to reconstruct the relative location and state of an object consists of two parts, one which can be transmitted instantaneously, but cannot be used without the other, which can only be transmitted by conventional means at the speed of light or slower. This quantum ability alone combined with the other undoubted advantages of holographic information processing such as associate memory, redundancy, etc, could account for the observation that biological systems/brains have some operations that are astounding from the engineering, technical viewpoint. **Implicit in the hypothesis of the central challenge, this duality of communication capabilities implies that information processing in biological systems occurs in tandem at both the**

traditional classical and quantum levels. ***By this means, for example, when living cells replicate, each cell would be fully aware of its sister cells relative locations and states, explaining how such replication is able to construct the morphology and dynamics of an embryo subject to the master blueprint of their DNA, and how, when the embryo is complete, self-repair/immune system is brought into play to restore such locations and states,ie health.***

5. DEVELOPMENT HISTORY

Physics ;begins with:-

a) Kepler, who first used a concept of phase in his analysis of Tycho Brahe's data to deduce his laws. A recent formalisation of this analysis has been successfully applied to functional magnetic resonance tomography (MRI) by Schempp[18],

b) Huygens' conceptualization of the behaviour of wavefronts -Huygens' principle- used extensively in classical wave physics, and by Feynman as the basis of his sum of histories approach to quantum physics. It was formalised in various forms by Jessel[19]- leading him to the discovery of anti-sound- and by Resconi and Jessel[20].

Both concepts are central to quantum physics, as applied by de Broglie, Schrodinger ,Bohm, Feynman and now Berry(in relation to the geometric phase/topological effects), as are,

c) Lie transformations, a culmination of geometrical invariance theory, applied as central to perception, cognition and brain structure by Hoffman[21],

d) holography, the phenomenon of full wavefront reconstruction, discovered by Gabor[17], together with a new computing principle[22], which defines a universal non-linear analogue filter, simulator and predictor which optimizes itself by a learning process - a prototype of which was built at Imperial College in the late '50s. This principle was generalised (using Huygens' principle) by Resconi and Fatmi[23], and Fatmi, Jessel, Marcer and Resconi[24] (using Lie transformational systems, and Lie commutators, respectively); both papers had the support of Salem. They established the relation with topological effects, and the other various models of computation - Turing, Deutsch, etc, and

e) Hahn, a disciple of Bloch, who pioneered magnetic resonance scanning and Hounsfield, who devoted the latter part of his Nobel prize address to the future of MRI rather than X-ray tomography. A full theoretical approach to MRI has been developed by Schempp[7] based on harmonic analysis on the Heisenberg group with its applications to signal theory[14]. This nilpotent Lie group, able to model Huygens' principle, holography, etc ; its Lie algebra, Heisenberg uncertainty(as was realised by Weyl in 1928); and its nilmanifold, topological effects, etc., defines quantum holography. Based on the Lie commutator [q,p], it specifies a quantum model of Gabor's universal, analogue, non-linear, filter, simulator and predictor that optimizes itself by a learning process, employing the phenomenon of full wave front construction ie holography, uniting a),b),c),d)and e)..

Living systems :-

This is already a significant history of publication and scientific workers, who anticipated this paradigm shift, predicting that various aspects and properties of biological systems, anomalous to classical explanations, could only be accounted for quantum mechanically. These include, Schrodinger, Pauli, Jung, Bohm, Frohlich, Pribram, Umezara, Jibu, Yasue, Eccles, Josephson, Penrose, Hameroff, Ho, Hiley, Farre ,Grossing, Schempp, Mitchell, etc

6. CONCLUSION

This approach to wider perspectives is an established programme of research, guided by the understanding of the known workings of a production device, the fMRI tomographic machine, for which an extensive, validated, theoretical description, quantum holography, exists.

The continuation of this programme, in accordance with existing and new models (*in italics*) to validate or refute the central challenge proposed and various subgoals (in **bold**), would require minimum financial investment, manpower, and timescale (say five years). The result would be

i) the extent to which the new paradigm will change our understanding of science and technology in the new millennium, with

ii) the critical and immediate implications, this could have on the critical masses of various (possibly all) research, development, and higher education programmes.

A successful validation - the prestige of discovering the quantum physics underlying the origin and functioning of living systems, would constitute a quantum leap in mankind's understanding of Nature, of how brains work; how cognition/intelligence can be understood and are carried out; and understanding of biological evolved engineered solutions would lead to the rapid technological development of manmade ones !

7. REFERENCES

[1] Various Authors, Quantum information, New forms of computation and communication, Physics World-special issue-1998,vol. 11, 3, pp 33-57.

[2] Rice S. A., New Ideas for Guiding the Evolution of a Quantum System, Science, vol. 258, 16th Oct. 1992, pp 412-413.

[3] Dahleh M., Peirce A.P. and Rabitz H., Optimal control of uncertain quantum systems, Phys. Rev.A, vol. 42, no. 3, lst Aug., 1990, pp 1065-1079.

[4] Judson R.S. and Rabitz H., Teaching Lasers to Control Molecules, Phys. Rev. Letts. vol. 68, no. 10, 9th March, 1992, pp 1500-1503.

[5] Bouwmeester D., et al. Experimental quantum teleportation, Nature,1997, 11th December, vol.390, pp 575-579.

[6] Sudbery T., The fastest way from A to B, Nature 1997, 11th December, vol.390, pp 551-552.

[7] Schempp W., Magnetic Resonance Imaging: Mathematical Foundations and Applications, John Wiley, NY, 1997.

[8] Marcer P.J. and Schempp W., A mathematically specified template for DNA and the genetic code in terms of the physically realisable processes of quantum holography, 1996, March 9th, Proc. the Greenwich symposium on Living Computers, eds. Fedorec A.M., Marcer P.J., pp 45-62.

[9] Marcer P.J. and Schempp W., The model of the prokaryote cell as an anticipatory system working by quantum holography, Proc. 1st Intern. Conf. Computing Anticipatory Systems, Belgium,Aug. 11-15, 1997, ed. Dubois D. pp 307-313.

[10] Marcer P.J. and Schempp W., The Brain as a conscious system, Intern. Journal General Systems, 1998, in press.

[11] Marcer P.J. and Schempp W., Model of the Neuron Working by Quantum Holography, Informatica, vol. 21, 1997, pp 519-534.

[12] Schempp W., Quantum holography and neurocomputer architectures, Journal of Mathematical Imaging and Vision, vol. 2, 1992, pp 279-326.

[13] Schempp W., Bohr's indeterminacy principle in quantum holography,self-adaptive

neural network architectures, cortical self-organization,molecular computers, magnetic resonance imaging and solitonic nanotechnology, Nanobiology, 1993, vol.2, pp 109-164.
[14] Schempp W., Harmonic Analysis on the Heisenberg group with applications in signal theory. Pitman Notes in Mathematics Series 14. Longman Scientific and Technical, London, 1986.
[15] Fink M., Time Reversed Acoustics, Physics Today, 1997, March, pp 34-40.
[16] de Duve C., A Guided Tour of the Living Cell, Scientific American Library, W. H. Freeman, New York, 1984, two volumes ; in particular, vol. 2, pp 250, pp 257.
[17] Gabor D. and Goss W.P., Interference Microscope with Total Wavefront Reconstruction, J. Opt.Soc.America, 1966, vol.56, 7, pp 849-858.
[18] Schempp W., Mysterium Cerebrotopographicum, (early private communication).
[19] Jessel M., Une formulation analytique du principe de Huygens, Comptes Rendues, 1954, vol 239, 8th Decembre, pp 1599-1600.
[20] Resconi G. and Jessel M., A General System Logical Theory, Intern. J. General Systems,1986, vol.12, pp 159-182.
[21] Hoffman W.C., The Visual Cortex is a contact bundle, Appl. Maths. and Comp.,1989,vol 32, pp 137-167; an important late paper in an extensive history of publications on this subject.
[22] Gabor D. et al., A Universal non-linear filter, predictor and simulator which optimizes itself by a learning process, Proc. IEE,1960, 108B, pp 422-438.
[23] Fatmi H. and Resconi G., A new computing principle, Il Nuovo Cimento, 1988, vol. 101B, 2, pp 239-242
[24] Fatmi H. et al., Theory of Cybernetic and Intelligent Machine based on Lie commutators, Intern.J. General Systems, 1990, vol. 16, pp 123-164.
[25] Jones J.A. et al.,Nature, 1998,393, 28th May,pp 344-346.

The Jigsaw, the Elephant and the Lighthouse.

Peter Marcer
53 Old Vicarage Green, Keynsham, BS31 2DH, UK.

1. INTRODUCTION.

The central problem of quantum computing concerns how to manipulate quantum linear superpositions, and in particular, but in my opinion perhaps more difficult, how best to control quantum measurement in order to correctly extract the output of the resulting quantum computation.

Significant light has been shed on these areas in relation to the quantum control of chemical reactions. In particular, Shapiro [1998] cites holographic methodology as the linear and stable means to calculate the phases, which determine such controls, in contrast to non-linear techniques which are unstable.

Astonishingly, however, in relation to the field of nuclear resonance spectroscopy began by Felix Bloch, Nobel Laureate and student of Heisenberg, the basic nature of such control has already been known for many years, since the work in 1950 of Hahn [1990]. This knowledge has been extensively refined and applied with great success to the workings of (nuclear) magnetic resonance imaging (fMRI) technology and machines, which are now in production use throughout the world for routine medical diagnosis. The mathematical specialist on the workings of such machines and how best to control them to best extract more highly resolved body and brain scan images faster than ever before, is Walter Schempp [1986, 1998]. His [1992] mathematical presentation, called quantum holography, validated for the case of fMRI, describes the ability, applicable to any kind of wave, to create, write, read, erase, filter, transmit and teleport holographic information, ie holograms, utilising quantum linear superposition. Holograms also incorporate their own natural information redundancy or error correction encoding. Quantum holography is therefore, as I shall attempt to show by independent mathematical means, the solution to this most difficult problem. In particular, quantum holography is notable for an explicit mathematical representation of the process of the quantum measurement, which is essentially absent or implicit in other approaches using the standard quantum mechanical formalism. This is not because of any inadequacy in the existing quantum mechanical axioms, but because the current approaches fail to reveal quantum measurement's extremely rich mathematical structure, the skeleton of which I shall now expose with the aid of three metaphors.

It is to be noted that in both quantum control of chemical reactions, and Schempp's (ensemble) quantum holography, it is phase, the new class of quantum observable [Resta, 1997], that is the essential quantity of physical significance.

2. THE THREE METAPHORS.

The metaphors are 1) the jigsaw, 2) the elephant, and 3) the lighthouse.

The mathematics to be presented, says that : -

1) the unique perspectives, with respect to the observed from the viewpoint of an observer, together make up a whole picture. These perspectives are therefore the pieces that make up the jigsaw.

2) such observers may be compared to the three blind philosophers taken to feel an elephant, who respectively describe the tail, a leg, and a tusk; leading each to an

observation apparently at odds with the others, but which, in fact, is a consistent aspect of the whole - an elephant, and

3) now consider an observer walking on the seashore on a dark night with thick, low cloud, who sees only a faint light reflected from the surface of the waves. Such an interference of light with water waves is a hologram ; an encoding, which could be decoded, if one could control the source of the water waves, so the surface of the sea becomes flat; that is, like a mirror in which one sees the image of source of the light waves -a lighthouse- the source of the illumination in this case! [Note,1] This metaphor is almost an exact analogy for the holographic encoding and subsequent decoding of a hologram, or for the encryption/decryption of an image or a message, where the original wave interference is, in fact, an exponential map or 'disordering', and so is seen as random ; although clearly in fact, it is not !!

This raises the question that, if now it is possible to exercise quantum control so as to "sculpt the wave function" performing bond-selective chemistry [Schleich,1999] and extract brain slice images by ensemble quantum holography, is there such a thing as true randomness, or is there simply lack of knowledge ie uncertainty, or as is essential for quantum entanglement, indistinguishability?

3. THE BONES OF THE MATHEMATICAL SUBSTANTIATION

The commutative diagram (A) expresses the syllogistic representation of the fundamental spectral theorem of Hilbert and Von Neumann, setting out the relationships/morphisms in the Hilbert space, between the vectors f, their corresponding self-adjoint operators F, measure functions W and rotations R.

Diagram A

THE FUNDAMENTAL SPECTRAL THEOREM
OF HILBERT and VON NEUMANN

$$\begin{array}{ccc} f & \xrightarrow{\quad R \quad} & R(f) \\ F \big\downarrow & & \big\downarrow W \\ F(f) & \xrightarrow{\quad R \quad} & RF(f) = WR(f) \end{array}$$

1. The Lie and therefore implicitly topological group presentation of the diagram shows that :-

a)for each R(f), an inverse rotation R^{-1} exists such that $F=R^{-1}WR$, ie that a diagonalization is possible for every F(f). Thus, this syllogism constitutes a model of observation/measurement and, indeed, quantum computation, since the vectors f can be put into orthogonal form as a quantum linear superposition and constitute a mapping onto the integers,

b) since in Lie transformational theory, the natural Lie diffeomorphism or differentiable mapping (morphism) with a differentiable inverse, is the exponential map (exp:) and so its inverse is the logarithmic map (log:), and it has been shown [Bowden,1996; Manthey,1996] that the linked chains of relationships between the corresponding homologies and cohomologies, as set out in figure (B), exists in

Diagram B.[Note,2] with acknowledgement to Keith Bowden.

$$
\begin{array}{ccccccccccc}
 & & \partial & & \partial & & \partial & & \partial & & \\
\ldots\ldots 4 & \rightarrow & 3 & \rightarrow & 2 & \rightarrow & 1 & \rightarrow & 0 & & \text{(dimension)} \\
op\downarrow\uparrow & & op\downarrow\uparrow & & op\downarrow\uparrow & & op\downarrow\uparrow & & op\downarrow\uparrow & & \\
\ldots\ldots 4 & \leftarrow & 3 & \leftarrow & 2 & \leftarrow & 1 & \leftarrow & 0 & & \text{(dimension)} \\
 & & \delta & & \delta & & \delta & & \delta & &
\end{array}
$$

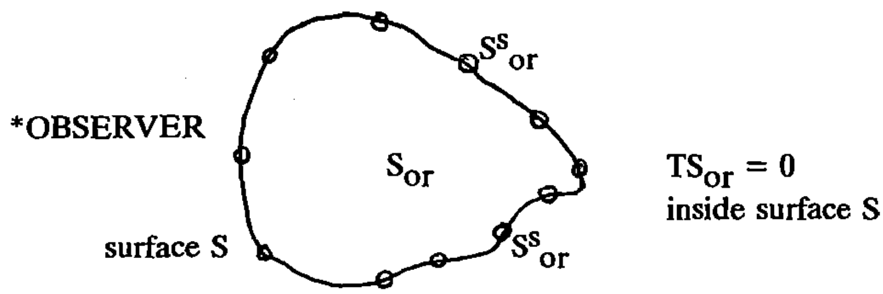
quantum mechanics. Thus the operators ∂ and δ are seen, in the Lie context (A), as generalised differential and integral operators, respectively, where for every appropriate homology, there is a corresponding cohomology with substantial structure in between ; (and where similarly, it might be postulated, since for every Lie group there exists a corresponding Lie algebra, that for every appropriate homotopy, there will a corresponding cohomotopy ?), and

c) this relational presentation in terms of the field of vectors f, would also imply by category theory, that there exists a dual interpretation in terms of objects. Diagram (C) is such an interpretation -the relational syllogism of Huygens' principle of secondary sources for the generalised propagation of a wave- discovered by Jessel and Resconi [1986], where the objects are the sources of the field. [Note,3]

Diagram C.

$$
\begin{array}{ccc}
F = F(0) & \xrightarrow{\quad OP \quad} & S_{or} = OPF(0) \\
T \big\downarrow & & \big\downarrow \\
F(t) = TF(0) & \xrightarrow{\quad OP \quad} & S^{s}_{or} = (OPT\text{-}TOP)F(0) \\
 & & = (OP,T)F(0)
\end{array}
$$

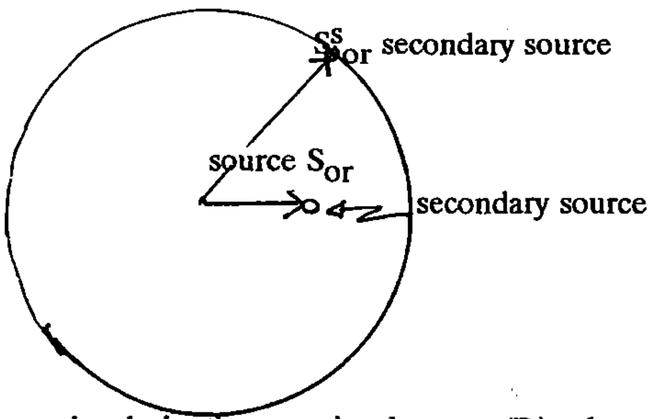
Example.



(C) expresses a field F in terms of its source S_{or} through the operator OP, such that S_{or} = OPF(0) and F(0) = $OP^{-1}S_{or}$ so that OP^{-1} exists, and relates the propagation of the field F(t) at time t to the formation of secondary sources S^{s}_{or} on a surface S, via an operator T. Such that, for example, TS_{or} is zero inside S, and in this case, therefore, from the point of view of an observer outside S, it is possible to substitute the source S_{or} by secondary sources S^{s}_{or} on S. Thus T is a Heaviside operator, equivalent to the singular Green's function (Schwartz distribution) that permits the corresponding description of the same phenomenon by means an integral formula. For example; T in three dimensions, corresponds to the Dirac delta function ; Feynman's sum of histories approach is such an integral formulation of quantum mechanics grounded on Huygens' principle; and in the lighthouse metaphor, T

represents the means to describe the calming of the sea within a boundary S, or in terms of its Heaviside dual, to restore its wave motion, ie to describe the performance of the necessary holographic decoding and encoding respectively, of the image of the lighthouse, which is the source of the illumination. [Note,4]
Thus the dual diagrams (A) and (C), where three dimensional space and time are implicit and explicit respectively, are seen to correspond within the standard quantum formalism to Heisenberg and Schrodinger (wave) formulations respectively, revealing what the standard formalism conceals ie the relationships of diagram (B); namely the hierarchical twin 'anti-parallel' chaining structure and the twisted isomorphisms that link them together.
2. An interpretation of diagram (A) is that it provides a unique global mathematical perspective from which to view all the other perspectives F(f) (defined on the quantum mechanical ray space of vectors f, figure (D)), as self similar and as constituting a signal/measure space in accordance with the measure functions WR(f). These perspectives (exemplified in the metaphor of jigsaw and elephant) are implicitly geometric, because of their implicit mathematical Lie presentation in the commutative diagram(A). Such an implicit geometric presentation can be made explicit, by introducing the three dimensional representation of the nilpotent Heisenberg Lie group, the Lie algebra of which is an expression of the Heisenberg uncertainty principle as was known to Hermann Weyl in 1928; the Heisenberg group is the basis of Schempp's formulation of quantum holography.
Figure D.

"RAY SPACE"



These perspectives -signals in the ray signal space (D)- therefore determine a Lie group invariance, appropriate to the observed from the perspective of an implicit observer. This can therefore be hypothesized as Lorentz invariance, which is known to hold universally in quantum mechanics with respect to the transmission of signals.
The mathematical description of diagram (A) applied to figure(D), thus constitutes a universal relativistic quantum holographic model of the theory -quantum mechanics- defined on the ray signal space; a computational model, which is a universal model of the quantum measurement process and is quantum computer constuctor universal , since it :-

a) applies for all vectors f in the Hilbert space
b) is Lorentz invariant
c) concerns self-similarity with respect to the perspectives F(f), which are implicitly geometric and known from Schempp's work to have an explicit geometric presentation in accordance with the Heisenberg nilpotent Lie group in terms of three parameters

x,y,z which are spatial measures. But a prototypical holographic machine for the control of quantum measurement as described by Schempp [1998], the (nuclear) magnetic tomographic resonance imaging machine already exists ! It has an interface to existing digital computer technology by means of which input and output from its quantum co-processor, are performed for the purposes of specifying the desired brain/body slice images required and their display, respectively ! At this interface, the output, a holographic diffraction pattern (hologram) obtained by spin echo techniques, is converted into the required image by fast Fourier transform techniques using the digital processor. And as will now be shown further, quantum holography as quantum measurement embraces the whole of quantum theory.

4. THE NATURE OF THE UNIVERSE OF DISCOURSE

From the ray signal space (D), it can be seen that the rotations R, diagram(A), can be used to define field automorphisms, which map the ray space onto itself, so the rays/signals of the space are under permutation ie concern the infinite dimensional permutation group, which is, of course, an exponential map. This introduces no infinities, because as diagram (A) shows an inverse transformation always exists, so that quantum holography is without renormalization problems. Thus these automorphisms, which, when normalised, concern unit rays having a common source S_{Or} at their centre, and secondary sources $S^{S}{}_{Or}$ at the unit ray ends :-

1) are clearly all self-similar,

2) cover the full spectrum of quantum mechanical behaviours by means of the infinite dimensional permutation group,

3) concern vectors f in the Hilbert space in quantum linear superposition, and signals, which derive from a common source, and thus correspond to entangled or squeezed states, such that each rotation R/automorphism converts one quantum linear superposition into another in a single step - a process exemplifying massive quantum parallelism,

4) moreover in the case of unit rays/signals ie normalization, the Lie diffeomorphism concerns exp: $0 = 1$, and the inverse mapping log: $1 = 0$. Thus, the automorphisms constitute a universal model of a theory in the language of sets $Z_2 = \{0,1\}$ and hence theorems of Erhlich [1989,1986] tell us, there exists a unique birthordering of these automorphisms. This can therefore be postulated as specifying the incremental evolution over time of this quantum universe (of discourse), where the initial signal automorphism of the discourse describes a initial resonance (or Big Bang). It is an evolution without end. Thus S_{Or}, in this case, is, because of the self-similarity, a quantum computer constuctor universal strange attractor!

This makes good sense in relation to quantum holography, a creation/annihilation model :-

i) as prior to each incremental massively parallel step n of the evolution by means of a rotation R_n $n=1,2,3,..\ \infty$, there will be the temporal and the spatial coherence necessary for holography, and S_{Or} will correspond to a squeezed signal state,

ii) at each massively parallel annihilation, the holographic source S_{Or} encodes all of holographic information appropriate to the evolution so far (for how else could the universe evolve by means of a unique birthordering?),

iii) at each creation, S_{Or} will be erased and substituted by its set of secondary sources $S^{S}{}_{Or}$, which, diagram (C) shows, perfectly simulate it (ie it is quantum computational

in the Deutsch sense),
iv) such simulations, by means of which the evolution proceeds, involve automorphic mappings, and so are phase conjugate adaptive resonances (ie holographic decoding /encodings of the universe onto itself), and
v) at each annihilation, the secondary sources S^s_{or} reconstitute their source S_{or}, incorporating the change happening in the previous evolutionary increment. This corresponds to the measurements/observations which took place with respect to time and space in that increment, ie the change of signal energies of those secondary sources.

5. UNIVERSE OF DISCOURSE AND COSMOLOGY?

Thus, although there is local time reversal symmetry, the universe of discourse describes a model cosmology with global time reversal asymmetry (ie, an arrow of time), as well as the fixed past and ever changing present such as we all experience. In terms of exp: $0 \rightarrow 1$ and log: $1 \rightarrow 0$, which correspond in logic to EXOR, the evolution in the ray signal space proceeds by shift register action on some topological nilmanifold, which describes the source of the signal field S_{or}. Furthermore, such a ray space model is irreducible, ie the simplest possible one, where a single ray/signal links each secondary source S^s_{or} to the source S_{or}. In Schempp's theory, this nilmanifold is that of the Heisenberg Lie group G, where, in accordance with the infinite but denumerable natural of the ray space, the signal field is described by means of infinite dimensional irreducible unitary linear representations U_v of the Schrodinger type, of G, unique up to an unitary isomorphism. Each element v of this class of representations realises $(U_v, L^2(\mathbf{R}))$ by means of quantum linear superposition, where v is frequency, **R** is the real line, and signal representations acting on the Hilbert space $H = L^2(\mathbf{R})$ concern its Schwarz space $S(\mathbf{R})$.
The act of self measurement in this model quantum holographic cosmology ie the normalization of amplitudes, thus makes phase the information carrier of physical significance. The Berry [1988]/geometric phase[Anandan,1992] is therefore of fundamental importance. It contains information about how long, where in three dimensional space, and what quantum states, the quantum system has visited, ie an historical record of the system's behaviour, which is, of course, essential to the birthordering that has already been proposed. The model cosmology is a unique automorphic birthordering or evolution, where processes are phase-locked together by means of frequency (ie think of gears, [Note,5]), both to the cosmos as a whole, and at many other scales throughout this evolution.
It is also possible from Berry's work to give the cosmological eigenvalues, since Berry [1986] has shown that there exists an unknown dynamic system, Hamiltonian H, with time reversal asymmetry, the eigenvalues of which are the imaginary parts of the zeros of the Riemann Zeta function, the real parts of which have all long been hypothesized to occur on the real line $x = 1/2$; a time asymmetric system, which from the above arguments can be hypothesized to be unique. It is also noteworthy that the Zeta function determines the distribution of the prime numbers, which are known to be linearly independent and approximate to an exponential map. It can be further hypothesized therefore that these eigenvalues determine the phase-locking appropriate to fermionic states ie $x = 1/2$, of fundamental importance in quantum computation with respect to quantum entanglement, such that the cosmos is, at every

stage of its evolution, a unique quantum whole, in accordance with Pauli's exclusion principle. Thus, Wittgenstein's principle holds, so that for each fact ie observation /measuremet/perspective, there is necessarily only one proposition that answers to it, and the sense of a proposition cannot be expressed except by repeating it. Thus the sense of computation or measurement, as used here, is to engage in producing a knowledge of facts; and once again we return to the metaphor of the jigsaw and the elephant!

6. SOME ADDITIONAL CONCLUSIONS

The postulated cosmology, is also in good accord :-

1)with Deutsch's original [1985] paper defining the universal quantum computer. Here programs admit all those defined in terms of all the ordinary Turing operations and just eight further operations. These are the unitary transformations confined to a single two-dimensional Hilbert space, the state space of a single bit (as a polarization or a spin 1/2), and

2) with Schempp's description of quantum holography as spin 1/2 choreography, which, in relation to fMRI, describes the controlled resonant coupling of electromagnetic fields with the quantum spin populations of the body's soft tissues, where the body slice image is extracted as a holographic diffraction pattern, figures A,B;C,D;E,F pages 13;14;22 respectively see [Schempp, in press]

Thus, this model cosmology in evolution is metaphorically seen as a self -choreographed dance of coherent energy, as for example is the understanding portaited by the Danse of Shiva in the cosmology of the Hindu religion; or, as the unique birthorder automorphism, the Way in the Chinese philosophy of the Tao.

3) In such a self measuring, self organizing system, a unique unobservable subsystem S_{or} is to be expected, since the measurement process or evolution needs a measurement standard against which to make each measurement, and in a self measuring system this clearly cannot be measured itself, because there is nothing to measure it against. Its unobservability also allows the secondary sources $S^{s}{}_{or}$ in the decoding phase, to be quantum teleported into existence from the source S_{or}, such that the incremental evolution appears continuous!

4) The natural Lie diffeomorphism ie exponential map (exp:) and the permutation group, say that in each domain of discourse, the usual statistical, statistical mechanical, and thermodynamical properties hold ie that is the Second Law of Thermodynamics holds, or will appear to hold, until the inverse mapping or laws which govern that domain are discovered/decrypted! For as shown here, (A) is such representation of natural laws, which give rise to this appearance, and consider

Diagram E

$$\begin{array}{ccc} S(0) & \xrightarrow{\quad OP \quad} & \sigma^{S} \\ T_1 \downarrow & & \downarrow T_2 \\ S(t) = T_1 S(0) & \xrightarrow{\quad OP \quad} & T_2\sigma^{S} = S(0)\partial T_1/\partial t + J^{S}.\nabla T_1 \end{array}$$

where $T_2\sigma^{S}$ is the new entropy production required to obtain the new behaviour S(t) and may be calculated in view of the fact that the entropy in an isolated system is governed by the continuity equation

$$\partial S/\partial t + \nabla . J^S = \sigma^S = OPS(0)$$

where $J^S = Sv$ is the flux of the entropy over a surface S (ie a hole in a larger surface, so that the diagram specifies the workings of a Maxwell demon). That is, in this syllogistic diagram, exactly analogous to (C), expressed however in terms of entropy density S, the entropy production σ^S can be used to create or destroy states in a thermodynamic system or machine, via say, a molecular or ionic field, so as to produce a new behaviour S(t) at time t, where again, $S(t) = T_1 S(0)$, and T_1 is again a Heaviside operator. That is to say, entropy behaves as an information metric, as well as as a measure of disorder, so that the diagram formalises the concept known as "negentropy"! This says that any apparent statistical randomness observed derives from the physical laws of a grand unification, and concerns a lack on knowledge on the part of observers. These laws specify a stochastic, self organized, optimally controlled evolution of an uncertain quantum cosmological whole or cosmos, rather than an increasingly disordered/random one.

5) Similarly, because the model cosmology evolves from an initial state to which it can never return, it is clear that the Third Law of Thermodynamics holds ; and the existence of a unique if unobservable Hamiltonian for the model cosmology implies that the First Law of Thermodynamics holds. [Note,6]

REFERENCES

Shapiro M. (1998) "Quantum State Holography" Physics Review Letters,80,7, 1418-1421 see also Rice S. A. (1992) "New Ideas for Guiding the Evolution of a Quantum System" Science,258,16th October, 412-413.

Hahn E. L. (1990)"NMR and MRI in retrospect" Philosophical Transactions of the Royal Society of London, A,333, 403-411.

Schempp W. (1998) Magnetic Resonance Imaging: Mathematical Foundations and Applications (Wiley-Liss, John Wiley and Sons, New York)

Schempp W. (1986) Harmonic Analysis on the Heisenberg Group with applications in signal theory: Pitman Notes in Mathematics Series 14 (Longman Scientific and Technical, London)

Resta R. (1997) "Polarization as a Berry Phase" Europhysics News, 28, page 19 "The Berry Phase"

Note 1. The author wishes to thank Michael Brown for this metaphor.

Bowden K. (1996) "Huygen's Principle, Physics and Computers or On General Physical Systems Theory II", Proceedings, 18th Annual (September, Cambridge) Meeting ANPA, ed. Etter T.L ; or see International Journal General Systems (1990) "On General Physical Systems Theories" 18, 61-79 and (1998) 27(1-3)...

Manthey M. (1996) "Distributed Computation, the Twisted Isomorphism and Autopoiesis" also Proceeding 18th ANPA meeting, 139-164.

Note 2. This diagram not only suggests a potential constructive model of evolution and learning, but also exhibits the basic characteristics of the structure of DNA, see Marcer P.J. and Schempp W. (1996) "A Mathematically specified template for DNA and the genetic code in terms of the physical realizable processes of quantum holography" Proceedings of the March 9th Greenwich Symposium on Living Computers, eds. Fedorec A. and Marcer P., 45-62.

Jessel M. and Resconi G. (1986) "A General System Logical Theory" International

Journal of General Systems, 12, 159-182.
Note 3. The essential complementarity of quantum phenomena ie wave/particle duality with fields as relationships/morphisms and particles as objects is essential mathematically (and philosophically!) so that quantum phenomena can be consistently categorised by means of functors and natural transformations !
Note 4. In actual quantum holography, the concept of the mirror cited in the lighthouse metaphor, is replaced by that of an active mirror working phase conjugately such that the object image and the object coincide. That is, a model quantum holographic eye/brain would see the incoming illumination from an object as outgoing illumination coincident with the 3D object in every particular with respect to the nature of that illumination. Thus, in this case, sight would be, as reality would be, a quantum holographic linear superposition; phase locked at some level of scale (ie the visible spectrum) to that reality. Thus a testable hypothesis, is that not only does quantum mechanics map onto reality, it is reality. Try out this hypothesis; just look out of your eyes and see the superposition for yourself !?
Ehrlich P. (1989) Proceedings Annual Meeting (January) American Mathematical Society, special session on surreal nimbers, invited lecture "Universally Extending Continua" see also N. Alling (1988) Foundations of Analysis over the Surreal Number Fields (North Holland, Amsterdam)
Ehrlich P. (1986) "The Absolute Arithmetic and Geometric Continua" Proceedings Philosophical Science Association,12, 237-246.
Deutsch D. (1985) "Quantum Theory, the Church-Turing Principle, and the universal quantum computer" Proceedings Royal Society of London, A400, 97-117.
Berry M. V. (1988) "The geometric phase" Scientific American, December, 26-32.
Anandan J. (1992) "The geometric phase" Nature, 360 (26), 307-313.
Note 5. In Schempp' theory such phase locking concerns, the generalised Heaviside operator appropriate to filtering, namely

$$\frac{\langle H_v(\psi, \Phi : x,y) \mid H_{v'}(\psi', \Phi' : x,y) \rangle}{\langle \psi \otimes \Phi \mid \psi' \otimes \Phi' \rangle} = 0,\ v \neq v' ; = 1\ v = v'$$

where H is the holographic transform of two wavelet mixing $\psi(t)dt \otimes \Phi(t')dt'$ where ψ, Φ are the wavefunctions. which concerns the probability of detecting a quantum, frequency v, within a unit area attached to point (x, y) in the hologram plane $R \oplus R$, and the above Heaviside operator holds for all ψ, ψ', Φ, Φ'
Example THE BASIS OF HEARING USING QUANTUM ACOUSTICS
Following the example of the lighthouse, quantum acoustics, where the quanta are air molecules vibrating at the acoustic frequencies of the observer (ie the hearer), it is a natural subdomain of the larger unified holographic field defined by the original syllogistic diagram A. Signals reaching the observer - now, the hearer - are therefore acoustic holograms, carrying encoded information including where the sources of the acoustic signals are located, such that with a suitable holophonic detection apparatus, this information can be decoded. Thus particularly sophisicated observers, ie bats for example, will be able to "see" the three dimensional world acoustically. Holophons may also contain 'symbolic' information in the form of a hierarchy of holophons ie letters/phonemes within words, words within sentences, sentences within paragraphs, etc, where such groupings are phase-locked together to convey a meaningful whole.
Berry M. V. (1986) "Riemann's Zeta Function: a model for quantum chaos?" Quantum Chaos and Statistical Nuclear Physics (Springer Lecture Notes in Physics no

263) eds. Seligman and Nishioka, Springer, Berlin, 1-17.
Schempp W. (in press) "Quantum Holography and Magnetic Resonance Tomography: an Ensemble Quantum Computing Approach" Proceedings of the Greenwich Symposium on the Frontiers of Computation, 14th March, 1998, eds. Fedorec A. and Marcer P.
Note 6. The evolution of the universe onto itself, described by the mapping of its Hamiltonian onto itself ie renormalization group theory (see K.G. Wilson (1982) "The 1982 Nobel Prize in Physics" Science, 218, 763-764), where new order parameters, such as magnetisation, concern the concept of stable and unstable fixed points, attractors, etc. Such stable fixed points are a "universal" representation of the different stable phases of matter and, for example, such an unstable or critical point, is that between steam and water, so that on all levels of scale, water vapour ie steam contains droplets of water and liquid water contains bubbles of steam from macroscopic,visible sizes down to the atomic, and so is a water/steam hologram!
Schleich P.W. (1999) "Sculpting a wavepacket" Nature,397, 21st January,207-208.

Link Theory –
from Computation to Quantum Physics
or
How to build a Universe using only things found lying around the math department

Part I: the Basics

Richard Shoup and Thomas Etter
Interval Research Corporation
Palo Alto, California
shoup@interval.com, tometter@mediacity.com
October, 1998

Abstract: Part I of this paper presents the basic ideas and mechanics of Link Theory, along with several examples from the realm of digital circuits, computation, and physics. In Part II, we show how the same principles can be used to derive the core laws of Quantum Physics, and how these laws are properties of the mathematics, not miracles, and not in any way derived from properties of physical matter.

0. INTRODUCTION

0.1 What is Link Theory?

Link Theory is essentially a calculus of composite abstract relations. As such, it is *not* itself a theory of computation or physics (or anything else), but a *tool* for creating and analyzing structures which represent digital circuits, physical situations including quantum physics experiments, and much more. Several examples are given below. Link Theory was developed primarily by Tom Etter beginning in 1996, but it has origins in much earlier work as well.

A *relation* is a set of mutually consistent *possibilities* among several variables. This set is in principle finite. Link Theory shows how to represent relations in tabular form, and how to compose relations into larger relations. A relation may also be viewed as a *constraint* on the set of all possible mutual values of several variables.

By *abstract* relation, we mean something which represents structural properties of a relation that don't depend on the kinds of things which are thus related. This means, for instance, that the structure of the Boolean relation (A = B AND C), as given by a table containing values 0 and 1, could just as well be given by a table with values "true" and "false", or "good" and "bad", or "Mars" and "Venus", or 0 and 1 reversed. The abstract AND relation is that which all such tables have in common.

1. LINK THEORY – THE BASICS

In this section, we describe the mechanics of Link Theory and give several examples of its use in representing digital circuits, solving algebraic equations, and studying computations in a computer operating on quantum mechanical principles.

Link Theory is a very simple calculus of composite relations in tabular form. It is closely related to tensor algebra of quantum theory, but can be used in a simpler form to represent digital circuits and other forms of computation. The great strength of Link Theory is its straightforward representation of *what is so* – or more to the point, *what is possible.*

1.1. Variables, domains, cases, case counts

To get started, we take the simple binary domain of digital logic. Each *variable* or signal can take on either of two *values*, which we designate as **0** and **1**. The value of a variable is seldom in any sense actual or definite in this calculus, but rather *the values represent possibilities.* It is from a simple enumeration of possibilities which we will build our universe.

Here is a set of common logic gates, each represented by a table consisting of *cases*, or rows, which enumerate all possible combinations of values among the variables indicated. Note that this representation is entirely symmetric, and no distinction is made between inputs and outputs. Unlike real electronic gates, no direction of action or causal relationship is implied or enforced here.

x	y	n
0	0	0
0	1	1
1	0	1
1	1	0

x	y	z	$n_\wedge$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

x	y	z	$n_\vee$
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Each table also includes an extra column labelled **n**, called the *case count* for that row. The case count may be thought of as the number of ways in which that combination of values can occur. In our initial examples, a case count will usually be 0 (impossible) or 1 (possible). A short inspection of the tables above will show that 1s in the case count columns correspond to the usual functions of these familiar gates. (It is important to distinguish the logical *values* designated **0** and **1** from the case counts 0 and 1, which are *numbers.* We could have used another nomenclature here, but both uses are standard and familiar.)

In general, case counts can be greater than 1, indicating hidden variables, and *less than 0,* as required to represent many quantum situations. The meaning of a *negative possibility* is indeed interesting to contemplate, but is mostly beyond the scope of this paper.

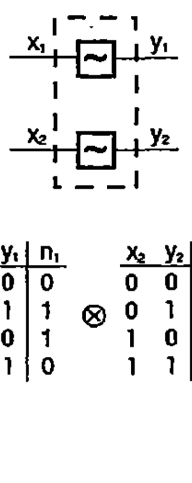
1.2. Relations

A Link Table may also be regarded as a *relation* or *constraint* on the joint values of its variables. Those cases which have a case count > 0 are permitted, that is *valid* under the constraint, those with case count = 0 are disallowed or impossible. A relation is inherently symmetrical, having no "input" or "output", unlike our usual conception of digital circuits. Relations are in general not functions in the mathematical sense, but all functions are relations. The bi- (multi-) directional character of relations can be used to great advantage in synthesizing and analyzing circuits.

1.3. Composing relations

Composing Link Tables is a simple matter of combinatorics -- accounting for all the possibilities. A composite *linked* relation or constraint is a *limitation* on the set of all the joint possibilities for the combined set of variables.

For example, if we consider two NOT gates taken together (but not connected), there are a total of four binary variables, and thus 16 total cases in their Cartesian product. But since both NOT constraints must be satisfied, only 4 cases are actually valid. *The case count in the composite table is derived by simply multiplying the case counts of the components for each case.*



x_1	y_1	n_1
0	0	0
0	1	1
1	0	1
1	1	0

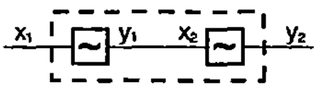
⊗

x_2	y_2	n_2
0	0	0
0	1	1
1	0	1
1	1	0

=>

x_1	y_1	x_2	y_2	n
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	0
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0
1	1	1	1	0

The diagram below again shows two NOTs, but with the addition of a connection, or *link*, between variables y_1 and x_2. In the composite table, these two variables must therefore have the same value in all cases, so 8 cases where this is not so have been crossed out. Note that only two of these were valid in the first place.



x_1	y_1	n_1
0	0	0
0	1	1
1	0	1
1	1	0

⊗

x_2	y_2	n_2
0	0	0
0	1	1
1	0	1
1	1	0

=>

x_1	y_2	n
0	0	0
0	1	0
0	0	0
0	1	0
0	0	0
0	1	1
0	0	1
0	1	0
1	0	0
1	1	1
1	0	1
1	1	0
1	0	0
1	1	0
1	0	0
1	1	0

=>

x_1	y_2	n
0	0	1
0	1	0
1	0	0
1	1	1

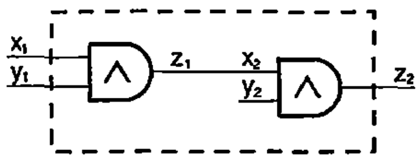
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} => \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Furthermore, we don't care at the moment what the values of y_1 and x_2 are, so we *hide* them, that is, we do not connect them outside the new composite circuit (dashed box). Variables x_1 and y_2 remain, and *are* visible and of interest. To get the overall result, we can simply erase the column associated with the hidden variable (the composite y_1- x_2), and *sum the case counts of like cases.* Four cases of the two variables remain, each with two instances in the composite table.

Summing, we get four cases, only two of which are possible. Apparently, linking two NOTs in this way yields the *identity relation,* since only the cases **0-0** and **1-1** have case counts = 1.

Note that we have included at the bottom of the figure an equivalent matrix multiplication, where the case counts are taken as matrix coefficients, each indexed by the corresponding logic values. Later we show the equivalence of Linking to matrix (more generally, tensor) multiplication.

Variables may be hidden or disregarded, as shown above, for various reasons. In the example below, we want to link the output of the first gate (z_1) to one of the inputs of the second gate (x_2), and then determine the relation which results between inputs x_1 and y_1 and the output z_2. We choose to hide or ignore both the linked variable z_1- x_2 and also the other input y_2. (An engineer would call y_2 a "don't care".) To do this, we link variables z_1 and x_2 in the AND gate tables as above, then hide (erase) the columns for z_1- x_2 and y_2, and combine like cases by summing their case counts.



x_1	y_1	z_1	$n_\wedge$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

$\oplus$

x_2	y_2	z_2	$n_\wedge$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

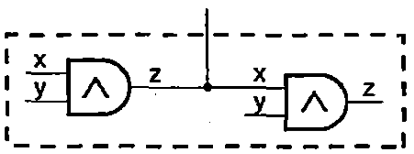
=>

x_1	y_1	z_2	n
0	0	0	2
0	0	1	0
0	1	0	2
0	1	1	0
1	0	0	2
1	0	1	0
1	1	0	1
1	1	1	1

1.4. Breaking a Link

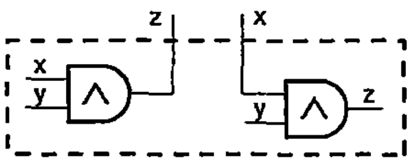
Links may also be ***broken***, and components which are thus separated are called *factored*. Below we show two AND gates linked much as before, but this time with only the linked variable **z-x** visible. (We have hidden all the others for this example. The reader may verify that case counts of 6 and 2 are the result.) We then break the link **z-x**, thus separating the variables **z** and **x** again. What remains is the cartesian product of the two tables taken together, with only **z** and **x** visible. (In physics terms, this is the density matrix relating the two variables) Finally, we separate the two components into two tables whose product is the composite table. This step is equivalent to factoring a matrix, as will be seen in Part II.

linked



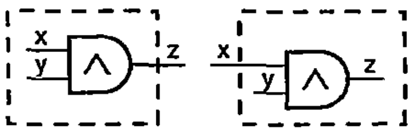
zx	n
0	6
1	2

cartesian product (link broken)



z	x	n
0	0	6
0	1	6
1	0	2
1	1	2

factored



z	n
0	3
1	1

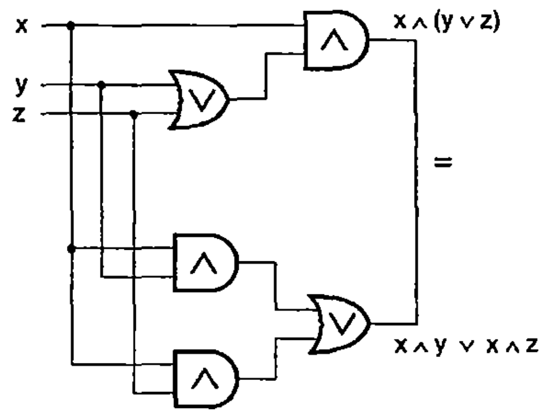
x	n
0	2
1	2

2. EXAMPLES AND EXTENSIONS

2.1 Example: Distributive Law

Below is somewhat larger example, again in the logic domain. By combining instances of the tables given above, and linking variables as previously described, we arrive at a simple "proof" of a Distributive Law of logic (AND distributes over OR). To represent the claimed equality, we simply connect (link) the two expressions together – something ordinary circuits (and most circuit tools) would not allow.

For every combination of values for x, y, and z, the composite relation is valid, as shown by the resulting case counts. If the circuit did not yield a tautology, the counts would indicate exactly those cases where the theorem failed.



x	y	z	n
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

2.2. Example: Arithmetic

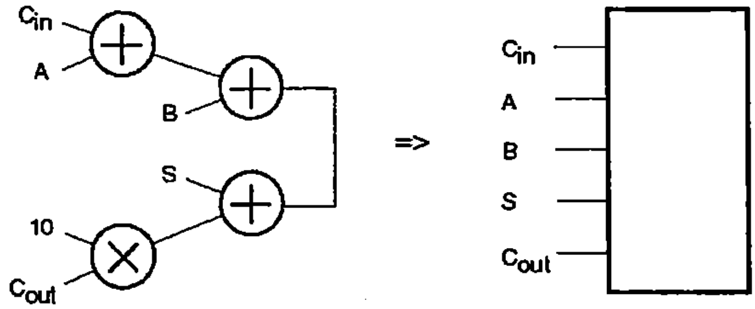
Link Theory has been successfully applied not only to logic domain problems, but also to numerical problems such as cryptarithms (shown here), finite-domain algebraic problems such as Diophantine (integer-solution) equations, and more. Here is a cryptarithm example usually considered in the domain of Artificial Intelligence. The problem is to find an assignment of single digits 0-9 to the letters such that the sum is satisfied. Each letter must correspond to a unique digit.

$$\begin{array}{r} \text{SEND} \\ \underline{+\text{MORE}} \\ \text{MONEY} \end{array} \quad \Rightarrow \quad \begin{array}{l} D + E = Y + C_1*10 \\ C_1 + N + R = E + C_2*10 \\ C_2 + E + O = N + C_3*10 \\ C_3 + S + M = O + C_4*10 \\ C_4 = M \end{array}$$

with $S, E, N, D, M, O, R, Y \in \{0\text{-}9\}$
uniquely, and $C_i \in \{0,1\}$

Also shown are the simultaneous equations implied by this sum, one equation per column. C_1 represents the carry out of the right-most column, and so on. We also assume that positional numbers do not begin with a leading 0, so M must equal 1.

To solve this problem using Link Theory, we first create a Link Diagram as shown below to represent each column constraint. Tables are used to represent the addition operator (only up to a sum of 19) and multiplication by 10. Reduction of this Link Diagram results in a single 5-column table, as shown.



$C_{in} + A + B = S + C_{out}*10$
for each column

By linking together 4 such composite tables according to the problem (linking E in column 1 to E in columns 2 and 3, for example), and applying the uniqueness constraints (S≠E, S≠N, S≠D, ... , E≠N, E≠D, ... , R≠Y) similarly, and reducing the overall composite Link Table, we are left with a table containing only one case, the solution:

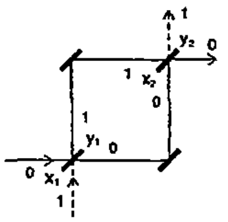
SEND		9567
+MORE	=>	+1085
MONEY		10652

2.3. Example: Quantum interference

In this example, we introduce a "gate" whose table includes a *negative case count.* (The deeper meaning of case counts < 0 is beyond the scope of this paper.) The diagram below represents an interferometer consisting of four mirrors -- two half-silvered and two fully-silvered. Four variables will be used to describe the possible paths of a photon through the apparatus.

A photon enters from the left (signified by variable $x_1 = 0$) or from below ($x_1 = 1$) and impinges on the first half-silvered mirror. Variable y_1 indicates the direction in which the photon leaves (upward, $y_1 = 1$, or to the right, $y_1 = 0$). The photon then bounces from a fully-silvered mirror and arrives at the second half-silvered mirror from the left ($x_2 = 1$) or from below ($x_2 = 0$), and exits upwards ($y_2 = 1$) or to the right ($y_2 = 0$). (For this example, we have simplified the actual physical situation and neglected complications introduced by the fully-reflecting mirrors.)

Using negative case counts, we can model a crucial aspect of the situation which actually occurs in nature. Tables for the two half-silvered mirrors are shown, and their the composite table, along with the corresponding matrix representation. From the matrices, it is easy to see that the half-silvered mirror acts as a 45° *rotation* on the photon state. In quantum physical terms, we have applied a *unitary transform* to the photon, and created a *superposition* of the two paths through the apparatus.



$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

x_1	y_1	n_1
0	0	1
0	1	1
1	0	-1
1	1	1

⊕

x_2	y_2	n_2
0	0	1
0	1	-1
1	0	1
1	1	1

=>

x_1	y_2	n
0	0	1
0	1	-1
0	0	1
0	1	1
1	0	-1
1	1	1
1	0	1
1	1	1

=>

x_1	y_2	n
0	0	2
0	1	0
1	0	0
1	1	2

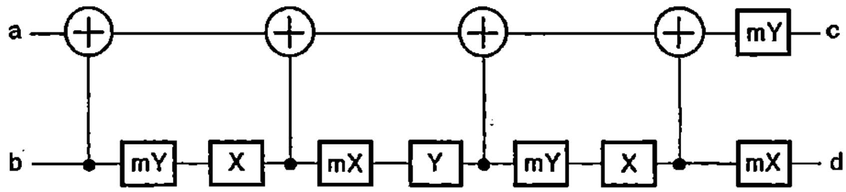
By hiding the linked variables y_1 and x_2, we are left with two cases for each of the four values of x_1 and y_2, similarly to the previous example. However, in two cases the negative values cancel with the positive values, and the result is 0, or impossibility -- in physical terms, *interference.* Thus from the composite table, we can see that any photon entering from the left *must* exit from the right, and a photon entering from below will necessarily exit upwards. (In Part II, we explore the connection to quantum physics in much greater detail.)

2.4. Quantum computation

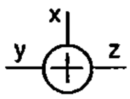
Recently, a number of research groups have been experimenting with computational mechanisms which operate in the quantum realm in hopes of building a *quantum computer.* A variable in a quantum computer can in principle represent a superposition of several states at once.

The following quantum computer circuit (or algorithm, equivalently) was suggested to us by Isaac Chuang of IBM Almaden Research Center. Using Link Tables, we calculated the composite transformation as described above, and achieved the identical result that standard quantum theory does. The reader may verify that compositing the given tables as shown in the diagram does give the result shown on the right.

Note that the imaginary case counts of $+i$ and $-i$ required by the given components can be accommodated by introducing two new variables linked in common among all elements, thus quadrupling the number of cases. This is equivalent to substituting a 2x2 spinor with real values for the imaginary matrix coefficients, as is sometimes done in quantum theory. This shows that imaginary case counts are not needed as a further extension to the theory, and in fact may be preventing a deeper understanding of the underlying structure of nature.



x	y	n_Y	n_{mY}	n_X	n_{mX}
0	0	1	1	1	1
0	1	1	-1	i	$-i$
1	0	-1	1	i	$-i$
1	1	1	1	1	1



x	y	z	n_{CNOT}
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

3. THE BRIDGE TO PHYSICS NOTATION

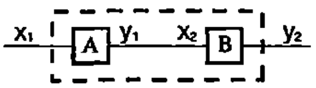
In a physicist's terms, the general binary pure state (in Dirac notation) $a|0\rangle + b|1\rangle$, corresponds to the Link Table

x	n
0	a
1	b

where a and b are scalars, and $|0\rangle$ and $|1\rangle$ represent the states corresponding to the logical values of **0** and **1**.

3.1. Equivalence to matrices/tensors

Returning to our simple two-component two-variable situation, if we substitute variables for the usual numerical case counts, and perform linking and reduction as before, it is easy to see that the result is the same as multiplying the corresponding matrices.



x_1	y_1	n_1
0	0	a_0
0	1	a_1
1	0	a_2
1	1	a_3

$\otimes$

x_2	y_2	n_2
0	0	b_0
0	1	b_1
1	0	b_2
1	1	b_3

=>

x_1	y_2	n
0	0	a_0b_0
0	1	a_0b_1
~~0~~	~~0~~	~~a_0b_2~~
~~0~~	~~1~~	~~a_0b_3~~
~~0~~	~~0~~	~~a_1b_0~~
~~0~~	~~1~~	~~a_1b_1~~
0	0	a_1b_2
0	1	a_1b_3
1	0	a_2b_0
1	1	a_2b_1
~~1~~	~~0~~	~~a_2b_2~~
~~1~~	~~1~~	~~a_2b_3~~
~~1~~	~~0~~	~~a_3b_0~~
~~1~~	~~1~~	~~a_3b_1~~
1	0	a_3b_2
1	1	a_3b_3

=>

x_1	y_2	n
0	0	$a_0b_0+a_1b_2$
0	1	$a_0b_1+a_1b_3$
1	0	$a_2b_0+a_3b_2$
1	1	$a_2b_1+a_3b_3$

$$\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix} \begin{bmatrix} b_0 & b_1 \\ b_2 & b_3 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \begin{bmatrix} a_0b_0+a_1b_2 & a_0b_1+a_1b_3 \\ a_2b_0+a_3b_2 & a_2b_1+a_3b_3 \end{bmatrix}$$

More generally, Link Tables can represent higher-dimensional matrices (tensors) and operations upon them. The advantage of Link Theory is a much simpler, more intuitive notation, and a more direct connection to the underlying combinatorics of the physical situation. With case counts > 0, we can represent logic circuits, sets, etc., while case counts < 0 are necessary to fully represent quantum phenomena.

4. SUMMARY AND IMPLICATIONS

We have shown how Link Theory, a simple accounting of possibilities, can be used to represent logic and idealized digital circuits, arithmetic problems, quantum experiments and much more. We also showed the equivalence of Link Theory to matrices and tensor algebra.

Link Theory is in fact a *general theory of structure.* Using Link Theory, we can show the equivalence and correspondence of theories in previously completely disparate realms, and *thus bring the theorems and the power of each to the others.*

In Part II, we show how the core laws of quantum mechanics can be derived from simple, entirely mathematical considerations.

5. REFERENCES AND RELATED WORK

Etter, Tom, and Noyes, H. Pierre (1998), "Process, System, Causality, and Quantum Mechanics: A Psychoanalysis of Animal Faith", Stanford Linear Accelerator Center Publication 7890. [www.slac.stanford.edu/pubs/slacpubs/7000/slac-pub-7890.html]

Russell, Bertrand (1959), *My Philosophical Development,* Routledge, London.

Spencer-Brown, George (1969), *Laws of Form,* George Allen and Unwin, London.

Parker-Rhodes, Frederick, *Indistinguishables* (unpublished manuscript).

Stein, Irving (1996), *The Concept of Object as Foundation of Physics,* Peter Lang, New York.

Bastin, T., and Kilmister, C.W. (1995), *Combinatorial Physics,* World Scientific, Singapore.

Gershenfeld, Neil, and Chuang, Isaac (1997), Bulk Spin Resonance Quantum Computation, Science, vol 275, p. 350.

Link Theory –
from Computation to Quantum Physics
or
How to build a Universe using only things found lying around the math department

Part II: the Miracle

Thomas Etter and Richard Shoup
Interval Research Corporation
Palo Alto, California
tometter@mediacity.com, shoup@interval.com
October, 1998

Abstract: Part I of this paper presents the basic ideas and mechanics of Link Theory, along with several examples from the realm of digital circuits, computation, and physics. In Part II, we show how the same principles can be used to derive the core laws of Quantum Physics, and how these laws are properties of the mathematics, not miracles, and not in any way derived from properties of physical matter.

0. INTRODUCTION

In the following sections, we discuss matrix and Link Table representations for states, projections, and transformations in quantum mechanics. We shall see how Link Theory leads to the core laws of quantum mechanics, and more important, to a generalization of these core laws that encompasses classical and other structures as well. By the core laws, we mean von Neumann's generalized statements of the Born probability rule, the average quantity rule and the Schroedinger equation. These three laws by no means exhaust quantum mechanics, since they make no mention of space, time, energy or any other specifically physical quantity. They do, however, encompass the essential novelty and strangeness of quantum mechanics, and are precisely that part of physics needed for the logical design of quantum computers.

The core laws will first be introduced without any Link Theory. Next, we will see how to derive them as general theorems of Link Theory. Finally, we look at quantum measurement or *quantization*, and show that it does not require any miracle or "collapse" of any wave function. Our treatment will necessarily be brief and informal; a fuller and more rigorous account can be found in (Etter and Noyes, 1998).

Somewhere in the distant past, arithmetic began as accounting. The need to tally cattle and other commodities led gradually to the concept of number and operations of addition, subtraction, and so on. Eventually, the idea of number and the calculating methods of arithmetic became abstract, separate from their applications. Similarly, Gaussian distributions were first derived from various physical situations, and only later abstracted into mathematics. Until Newton, many people considered the directions "up" and "down" to be immutable properties of space itself.

Today, our theories of quantum phenomena includes several new forms of mathematics which are considered, at least implicitly, to be properties of matter, belonging strictly to the domain of physics. Typically, the Quantum Core laws discussed below are introduced in a complicated way as *Laws of Physics* only at an advanced level of study.

It is the purpose of this work to show that another abstraction is possible: The Quantum Core turns out to be an unavoidable consequence of certain simple mathematical definitions in Link Theory, which is a general theory of abstract structure based on Russell and Whitehead's theory of relations (See Part I of this report for a discussion of the basics of Link Theory.)

In Step One, we will derive the "classical" case of the quantum core laws without using either Link Theory or physics, only common sense. Next, in Step Two we introduce the "miracle" needed in standard quantum theory to get the core laws in their full generality. Finally, we show how this miracle can be replaced by straightforward combinatorial analysis, *thereby factoring the core laws entirely out of physics.*

1. STEP ONE

Below is a statement of the core laws of quantum mechanics. P, Q, D, and T are matrices. We will see that these laws are just *common sense* (and contain no physics) for the special case when P, Q, and D commute. (We can think of this special case as describing the classical limit of quantum behavior.)

The probability rule: Prob(P,D) = trace(PD)

The average rule: Average(Q,D) = trace(QD)

The state transformation rule: $D' = T^{-1}DT$

Let us start with a set of objects classified by *states*, some *statements* about them, and numerical functions of these states called *quantities*.

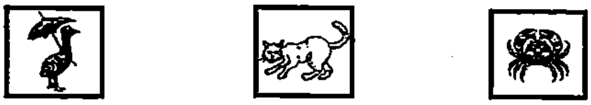
1.1 Classical states and statements

Suppose there are 100 children's blocks on the table – 60 birds, 10 cats, and 30 crabs.

60 birds 10 cats 30 crabs

The *state* of a block, i.e. Bird, Cat, or Crab, will be associated with a 3x3 matrix of 0's and 1's called the *state matrix* S. (The appropriateness of this representation will become clear later.) Each matrix contains a single 1 in the diagonal thus:

1,1 *Bird*	2,2 *Cat*	3,3 *Crab*
$\begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix}$	$\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix}$	$\begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix}$



Any *statement* about block X will also be assigned a matrix P called a *projection* that shows the states for which it is true. Here are some examples:

X has more than two legs	X lays eggs	X is a crab	X is anything
$\begin{pmatrix}0&0&0\\0&1&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$

1.2. Boolean logic as matrix algebra

Now we can perform simple logical operations using this representation. For example, the matrix of the statement (A AND B) is just the matrix of A multiplied by the matrix of B:

X has more than two legs AND X lays eggs — X is a crab

$$\begin{pmatrix}0&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix} = \begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix}$$

The matrix of statement (NOT A) is the identity matrix minus the matrix of A:

$$\underset{\text{X is NOT a crab}}{\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}} = \underset{\text{X is anything}}{\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}} - \underset{\text{X is a crab}}{\begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix}}$$

1.3. Density matrices and probabilities

The *density matrix* D shows the proportion of creatures in each state. Since there are 30 birds, 10 cats and 60 crabs in this case, we have

$$D = \begin{pmatrix}.3&0&0\\0&.1&0\\0&0&.6\end{pmatrix}.$$

1.4. The generalized Born probability rule

The *trace* of a matrix is defined as the sum of its diagonal elements, so

$$\text{trace}(D) = .3 + .1 + .6 = 1.0 .$$

The Born probability rule of quantum mechanics states that the probability of statement P about a collection having density matrix D is given by the trace of the product of P and D:

$$\text{Prob}(P,D) = \text{trace}(PD) .$$

Example:

$$\text{Prob(X is not a crab)} = \text{trace}\left(\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}.3&0&0\\0&.1&0\\0&0&.6\end{pmatrix}\right) = .4$$

1.5. Quantities

A *quantity* is defined as any numerical function q(S) of state S. It will be represented by a diagonal matrix Q that has q(S) in each S's diagonal place. Here is the matrix Q for the function "number of legs" :

$$Q = \begin{pmatrix}2&0&0\\0&4&0\\0&0&8\end{pmatrix}.$$

1.6. Averages

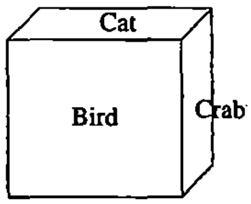
To find the *average* of a quantity in a collection, we multiply the quantity for each state by the proportion of members in that state and add the results. This is the same as multiplying the quantity matrix by the density matrix.

Definition: Average(Q,D) = trace(QD). So for this case,

$$\text{Average}(\mathbf{Q},\mathbf{D}) = \left[\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{pmatrix}\begin{pmatrix} .3 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & .6 \end{pmatrix}\right] = 5.8\,.$$

1.7. Transformations

So far we have been looking down on the blocks. Actually, all of the blocks are identical; one animal is on the left & right faces, another on the front & back faces, and the third on the top & bottom faces.



1.8. Permutations of matrices

In general, any linear transformation T of a matrix M can be written:

$$M' = T^{-1} M\, T$$

A *permutation* is a special case of a transformation, defined as a matrix with a single 1 in each row and each column, and 0's elsewhere. Here is the *cyclic* permutation T = [2, 3, 1] applied to D:

$$\underset{\mathbf{D'}}{\begin{pmatrix} .6 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & .1 \end{pmatrix}} = \underset{\mathbf{T^{-1}}}{\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}}\underset{\mathbf{D}}{\begin{pmatrix} .3 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & .6 \end{pmatrix}}\underset{\mathbf{T}}{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}}$$

In this example, the application of T has permuted birds to cats, cats to crabs, and crabs to birds.

1.9. The core laws

With just the above simple considerations, we have already seen in primitive form the three core laws of quantum mechanics, expressed as pure mathematics, without invoking any Physics at all:

The Born Probability Rule: Prob(P,D) = trace(PD)

The Average Rule: Average(Q,D) = trace(QD)

The Transformation Rule for states: $D' = T^{-1}DT$

Now let's make the simple generalizations and interpretations which show how these laws fully represent quantum situations, dynamics, and measurement.

2. STEP TWO - THE MIRACLE OF QUANTIZATION

The Standard Miracle that produces the quantum core is straightforward given the preliminaries above: Allow the transformation T to include *all rotations*, not just cyclic permutations.

More formally, we generalize to *closure under rotation*: The class of transformation matrices T is enlarged from the class of permutations of axes to the class generated by all rotations and reflections. This type of matrix is also known as a *unitary* transformation – one whose inverse is equal to its transpose. That is, $T^{-1} = T^*$.

[Cartoon by Sidney Harris:
"Then a miracle occurs..."
"I think you should be more explicit here in step two."]

The core laws are assumed to still hold. This requires that the P, Q, and D classes be enlarged to include all unitary transforms of their diagonal matrices as well.

All quantum weirdness (interference, EPR non-locality, etc.) resides in the quantum core. Note that, when viewed this way, the quantum core says nothing about space, time or matter.

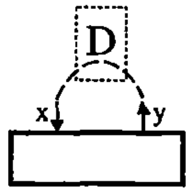
2.1. Step Two made explicit

Step 2 above makes no sense at all in terms of classical experience; It's a weird and seemingly ad-hoc intruder into the common-sense world. We'll now see how to completely represent it by simple case counting in Link Theory. The theory needed is only a slight extension of that described in Part I, in that we will now allow cases to count *negatively* as well as positively.

The crucial step is to *interpret density matrices as the matrices of broken links* (see Part I, section 1.5). This leads immediately to the Born Rule and Average Rule at every link, and to the Transformation Rule for the links on either side of any doubly-linked two-column table. The quantum form of these rules follows from the single additional assumption that every link is symmetrical in its two linked variables.

2.2. Breaking a link

Breaking a link between variables x and y defines a matrix which will be called the *link state* of x or of y. Link states defined in this way act as density matrices, thus D is the density matrix which would result if we broke the link between x and y:



2.3. Types of density matrix

Not all density matrices defined by breaking a link are *quantum states.* The defining condition for a quantum state is that the matrix is unchanged by reversing x and y. A matrix with this symmetry property is called *self-adjoint.*

A state is *pure* if x and y are independent. If a pure state is self-adjoint, then it is a *quantum pure state,* and thus x and y are the same vector. The (normalized) counts in vector x (or y) are what in physics are called the *amplitudes* of the quantum state.

In a *causal state* (which is usually pure), one vector is "white" (all counts equal), while the other vector carries all the information.

Quantum state: $D = D^*$.

Quantum pure state: $D = | v > < v |$. The vector v is the wave function.

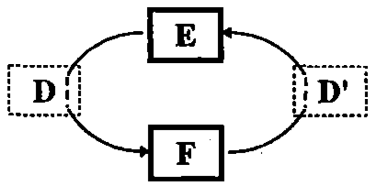
Causal pure state: $D = | p > < 1 |$,
where p is a probability distribution vector and 1 is the white vector.

Note that quantum and causal are only two ideal limiting cases among an infinite variety of state types.

2.4. The Born Rule and Average Rule

Breaking a quantum link is *counterfactual,* i.e., we never *actually* encounter the broken ends x and y. We always see x and y *linked,* as a single variable. Since their broken distributions are equal, the actual distribution on x (or y) is always the square of that distribution; this proves the well-known version of Born's Rule which says that probability is the square of amplitude. From this, the von Neumann generalization of Born's Rule and the Average Law can be derived for diagonal operators by the same common-sense reasoning that we used in Step 1, and extended to non-diagonal operators by the generalized Schroedinger equation (see below) together with the theorem that trace(AB) = trace(BA).

2.5. Transformations of state

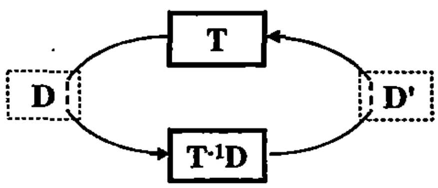


Suppose the tables E and F are linked as shown, with D and D' as link states on their respective linked variables. If the matrix of table E has an inverse, then $D' = E^{-1}DE$.

Proof: $D = EF$ (only D is broken), so $F = E^{-1}D$. But $D' = FE$ (only D' is broken), so by substitution $D' = E^{-1}DE$.

This is the core transformation law. If E transforms any quantum state D into another quantum state D', it can be shown that E must be unitary.

2.6. Example: cyclic permutation

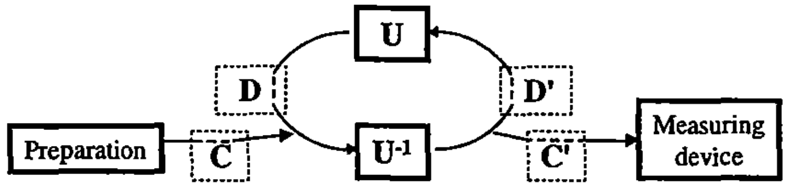


Let us return to the special case of permutation, where E is the permutation matrix T of section 1.8, and F is $T^{-1}D$, where D is the animal block density matrix of section 1.3. Then $D' = T^{-1}DT$.

Proof: The left link state is $TT^{-1}D = D$, while the right link state is $T^{-1}DT = D'$.

2.7. Quantum measurement

Now we are prepared to present the complete situation of quantum measurement.



U is any unitary transformation matrix. C is the (causal) density matrix of the *preparation*, a state vector which defines the initial state D of the quantum system to be measured. C' is the (causal) density matrix of the *measurement* which couples the final state D' to the measuring device. See (Etter and Noyes, 1998) for a more complete presentation.

There is no actual wave function to collapse -- it's counterfactual.

3. REFERENCES AND RELATED WORK

Etter, Tom, and Noyes, H. Pierre (1998), "Process, System, Causality, and Quantum Mechanics: A Psychoanalysis of Animal Faith", Stanford Linear Accelerator Center Publication 7890. [www.slac.stanford.edu/pubs/slacpubs/7000/slac-pub-7890.html]

Russell, Bertrand (1959), *My Philosophical Development*, Routledge, London.

Spencer-Brown, George (1969), *Laws of Form*, George Allen and Unwin, London.

Parker-Rhodes, Frederick, *Indistinguishables* (unpublished manuscript).

Stein, Irving (1996), *The Concept of Object as Foundation of Physics*, Peter Lang, New York.

Bastin, T., and Kilmister, C.W. (1995), *Combinatorial Physics*, World Scientific, Singapore.

Gershenfeld, Neil, and Chuang, Isaac (1997), Bulk Spin Resonance Quantum Computation, Science, vol 275, p. 350.

PHYSICS: LET'S GET DOWN TO BASICS

PETER ROWLANDS

Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford Street, P.O. Box 147, Liverpool, L69 3BX

ABSTRACT

A core level of basic information for physics is identified, based on an analysis of the characteristics of the parameters space, time, mass and charge. At this level, it is found that certain symmetries operate, which can be used to explain certain physical facts and even to derive new mathematical theorems. Applications are made to classical mechanics, electromagnetic theory and quantum mechanics.

1 A foundational level

Certain aspects of physics suggest the existence of a core level of basic information which is completely independent of any hypothesis or model-building. This information is concerned with the definition of the fundamental parameters of measurement and how they are structured. Here, we are not so much describing nature itself as specifying the characteristics of the simplest categories needed to make such a description possible. It is, of course, sometimes argued that physics should not necessarily be concerned with the simplest possible ideas because nature may well not be simple in principle, but this argument is based on a misconception. Physics as we know it has evolved because it has created a set of simple categories which have been successful in devising the human construct that we call the 'description of nature'. Whether or not such simplicity is truly characteristic of 'reality' is beside the point.

Now, the purpose of isolating a foundational level for physics would be to give a simple account of those important facts which are purely concerned with our own processes of measurement, and not with, say, the nature of matter or the structure of the universe. Such information, if isolated, could be dealt with more efficiently than if linked, unnecessarily, with more specific or more complicated theories, and could lead to the creation of wholly new sources of foundational information. In addition, if we pitch our foundational theories at a high level of sophistication, as we are often tempted to do, we cut ourselves off from understanding their origins. Many truly fundamental results have often turned out to be simple in principle, even when it has taken a sophisticated approach to find them. The simple bases are often found only after a prolonged struggle with more complicated ideas, but it is possible that we could discover some of them more easily by a direct analysis.

In attempting to reach the foundational level, then, we need to separate out the truly fundamental ideas from the mass of sophistications which inevitably accompany them, but the really basic ideas may not be particularly difficult to find. Obviously, for example, space and time are basic; it is impossible to conceive a theory of physics without space and time. This is not necessarily true for a combined space-time; combinations, by definition, are not basic. If we use space-time as basic, we lose sight of the separate identities of space and time, which are surely important at the fundamental level. It is certainly not true that in *combining* them we have explained one in terms of the other. We should really regard space-time, whether curved or otherwise, as a sophistication, as something to be discovered after we have investigated the separate identities of space and time.

Once we have selected space and time, there is only one other type of information that is likely to fit the description 'basic', for the fundamental bases of the whole of physics are undoubtedly the four known interactions: gravity; electromagnetism; and the strong and weak nuclear forces. Once we truly understand these, then we will also understand physics. But we already know something about them. We know, for example, that the *source terms* for gravity and the electromagnetic interaction are, respectively, mass and electric charge. We don't know the source terms for the weak and strong forces as precisely, but we know that such terms must exist; and, according to the rules of quantum electrodynamics, they should be more like electric charge than like mass. Also, although the three nongravitational forces are very different under normal conditions, the Grand Unified Theories of particle physics suggest that, under ideal conditions, they would be identical. It is reasonable, therefore, to assume that the differences between these forces are not basic, but a 'sophistication', to be explained in terms of the physical particles that actually exist, after we have established the basics. I have, therefore, found it convenient to describe the unknown sources for the weak and strong interactions as weak and strong 'charges', and to refer sometimes to the three nongravitational sources under the collective label 'charge'. (Something like this concept is actually used when we talk about the process of 'charge conjugation'.)

Thus, we have four basic parameters – space, time, mass and charge – and we can use such techniques as dimensional analysis to show that all other physical parameters arise from compounded versions of these elementary ones. We can also show that it is precisely these parameters which are assumed to be the elementary ones in the statement of the CPT theorem. It is perhaps slightly surprising that physicists have been so often prepared to tackle fundamental questions without taking proper account of these basic ideas. Although we can't hope to *analyse* really basic ideas, we can learn a great deal by setting one off against the other. It would surely be profitable to examine the properties of these parameters as closely as possible and look for patterns, or symmetries of one sort or another, that would help to clarify their meaning and uses. Symmetry has been such a powerful tool in understanding particle physics and the fundamental interactions that we have every reason to expect to find it here. We should, at any rate, examine the properties of our parameters to see whether or not it exists.

2 Conserved and nonconserved: mass and charge versus space and time

Perhaps the first thing that we notice when we closely examine space, time, mass and charge is that the last two are conserved quantities and the first two nonconserved. The conservation laws of mass and electric charge are among the most fundamental in physics, and we have every reason to believe that they are true without exception. Almost certainly, also, some type of conservation law applies to the other two types of charge and manifests itself in such properties of fundamental particles as lepton and baryon conservation. In addition, the conservation laws of mass and charge are not merely global, applying to the total amount of each quantity in the universe, but also *local*, applying to the amount of each quantity at a given place in a given time. It is as though each element of mass or charge had an *identity* which it retained throughout all interactions, subject only, in the case of charge, to its annihilation by an element with the opposite sign. We could, in principle, label each unique element with an identity tag which it would never lose. Mass and charge, thus, have identical conservation properties, apart from the fact that masses have no elements with opposite sign.

When we look at nonconservation, as manifested by space and time, we might at first imagine that it is merely the absence of conservation, but this is not so. Examination shows that it is the *exact opposite* of conservation and it is just as definite a property. Nonconservation is, in fact, one of the most interesting and important of all physical properties, and it manifests itself in many different ways. Thus, just as the elements of mass and charge have individual, specific and permanent identities, so those of space and time have no identity whatsoever, and this fact has to be incorporated directly into physics at all levels. We refer, for example, to the property of *translation symmetry* for both space and time. This means that every element of space and time is exactly like every other, and is not only indistinguishable in practice, but *must be stated to be indistinguishable* when we write down physical equations. And this translation symmetry is not an insignificant philosophical concept; it is responsible for two of the great laws of nature, for Noether's theorem shows us that the translation symmetry of time is precisely identical to the conservation of energy, and that the translation symmetry of space is precisely identical to the conservation of linear momentum.

Space, in addition, because it is three-dimensional, also has rotation symmetry; this means that there is no identity, either, for spatial *directions*. In addition to having no unique elements, space also lacks a unique set of dimensions. One direction in space is identical to any other; this is the fact that is responsible for space's famous *affine* structure, the infinite number of possible resolutions of a vector into dimensional components. Space rotation symmetry is a very important property and we will return to it later. Noether's theorem shows that it is exactly the same thing as the conservation of angular momentum.

The exactly opposite nature of conservation and nonconservation could be illustrated by expressing the identity or uniqueness properties of mass and charge in terms of 'translation' *a*symmetries. Translation asymmetry then means that one element of mass or charge cannot be 'translated to' (or exchanged for) any other within a system, however similar.

But translation and rotation symmetry are not the only manifestations of nonconservation in space and time. The whole of physics is based on defining systems in which conserved quantities remain fixed while nonconserved quantities vary absolutely. A conserved quantity can only be defined with respect to changes in a nonconserved quantity. In effect, we look at how mass and charge, and such quantities as energy, momentum, force or action remain constant, or zero, or a maximum or a minimum, because of the more fundamental requirements involving mass and charge, while the space and time coordinates alter arbitrarily. The alteration of space and time is expressed by describing them in terms of differentials. The very fact that we have based physics on differential equations and the definition of systems involving conservation requirements is an expression of the presence of both absolutely conserved and absolutely nonconserved terms in nature.

The absoluteness of the nonconservation properties is manifested in the *gauge invariance* used in both classical and quantum physics. In classical or quantum electrodynamics, electric and magnetic fields terms remain invariant under arbitrary changes in the vector and scalar potentials, or phase changes in the quantum mechanical wavefunction, brought about, essentially, by translations (or rotations) in the space and time coordinates. Gauge invariance tells us, in effect, that a system will remain conservative under arbitrary changes in the coordinates which do not produce changes in the values of conserved quantities such as charge, energy, momentum and angular momentum. In other words, we cannot know the absolute phase or value of potential because we cannot choose to fix values of coordinates which are subject to absolute and arbitrary change. Even more significantly, in the Yang-Mills principle used in particle physics, the arbitrary phase changes are specifically *local*, rather than global. Nonconservation, therefore, must be local in exactly the same way as conservation.

3 Real and imaginary: space and mass versus time and charge

Now, space and time are alike in their nonconservation, but we know that there must be fundamental differences between them; otherwise, they would be indistinguishable. One such distinction is evident in the very mathematical combination which produces four-dimensional space-time. This is the fact that, while Pythagorean addition produces positive values for the squares of the three spatial dimensions, the squared value of time becomes negative. A convenient way to represent time, then, is by an imaginary number, as in the Minkowski space-time 4-vector used in relativity. This, of course, does not make time 'imaginary' in itself; but it is important for us to ask why this particular 'trick' actually works. It is not really adequate to describe it as a 'convenience' without explaining why it is convenient. One interesting fact is that an imaginary representation would also make uniform velocity imaginary, while acceleration would remain real.

To try to get beyond the facile explanation that the trick is good because it works, we should see if we can learn anything relevant from the representation of mass and charge. Here, we have the intriguing fact, long known but never explained, that forces between like masses are attractive, whereas forces between like (electric) charges are repulsive; that is, the forces between like masses and like charges have opposite signs. Now, the force laws effectively square mass and charge terms, in the same way as space and time terms are squared in Pythagorean addition. Suppose, then, that we choose to represent charges by imaginary numbers and masses by real ones (figure 1). Such a procedure would be just as valid as using imaginary values of time when adding it to space.

$$F = -\frac{Gm_1m_2}{r^2}$$
$$F = -\frac{iq_1iq_2}{4\pi\varepsilon_0 r^2}$$

Figure 1. Symmetrical representation of Newton and Coulomb force laws using imaginary numbers for charges.

In addition, of course, the other two forces – the strong and weak interactions – are like the electromagnetic in being repulsive for like particles, and so the source terms for these forces would also presumably be defined by imaginary numbers. But the three types of source would have to be distinguished from each other in some way. And here we have a stroke of luck, for the mathematics required for such a situation is already available and has been well-known for a hundred and fifty years. This is the *quaternion* system, discovered by Hamilton in 1843, in which ***i***, ***j*** and ***k***, the three square roots of –1, are related by the formulae:

$$i^2 = j^2 = k^2 = ijk = -1 \,.$$

For historical reasons, quaternions became a proscribed concept at the end of the nineteenth century – they were not the kind of mathematics that respectable physicists used – and many people still think they are incredibly difficult or esoteric; but, in fact, they are remarkably easy to use, being just the reverse of the 4-vectors used in Minkowski space-time: three imaginary parts and one real (ordinary real numbers), as opposed to three real parts and one imaginary. But the real significance of quaternions is that they are unique. As Frobenius proved in 1878, no other extension of ordinary complex algebra involving imaginary dimensions is possible: if we require a dimensional imaginary algebra (as the source terms for the electromagnetic, strong and weak interactions suggest we might) then we have only one possible choice – an algebra based on one real part and three imaginary.[1]

Hamilton discovered the quaternions after finding that a system with two imaginary parts was impossible, and, almost immediately, he felt that he was on to the true explanation of 3-dimensional space, with time taking up the fourth or real part. By our analysis, it would be more convenient to apply them to the three imaginary components of charge, with mass taking up the real part. However, space and time would then become a three real- and one imaginary-part system by *symmetry*. In this sense, the three components of charge (say, *ie*, *js*, *kw*) could 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 (figure 2).

ix	*ie*
jy	*js*
kz	*kw*
it	*m*

Figure 2. Symmetrical representation of 4-vector space-time and quaternion mass-charge.

It is here that we can now return to the subject of the rotation symmetry of space. If charge, like space, is a three-dimensional parameter, then we need to investigate how the dimensions behave with respect to each other. Immediately, we should expect a difference from space, since charge is a conserved quantity. In fact, we should expect conservation in dimension as well as in quantity; in principle, charge should exhibit rotation *a*symmetry. That is, the sources of the electromagnetic, weak and strong interactions should be separately conserved, and incapable of interconversion. Immediately, this should tell us that the proton, which has a strong charge measured by its baryon number, cannot decay to products like the positron and neutral pion, which have none. Attention to basics here would require the separate conservation of the three charges to be built into Grand Unified Theories. Particle theorists have been puzzled as to why the proton does not decay; but basic reasoning suggests that there may be an answer. (The Weinberg-Salam unification of electromagnetic and weak forces is not, of course, affected because this theory is a statement of the identity of effect in the two interactions, under ideal conditions, not of identity of the sources; the three quaternion operators *i*, *j* and *k* are different sources, though identical in effect.) In addition, separate conservation laws would easily lead to baryon and lepton conservation, baryons being the only particles with strong, as well as weak, components, and leptons being the only particles with weak, but no strong, components.

In view of such advantages in applying an imaginary representation to the three types of charge, we may be inclined to ask if there is any further benefit; and, in fact, there is, for imaginary numbers have yet another important property. This is the fact that equal representation must be given to positive and negative values of imaginary quantities. Unlike real numbers, imaginary ones allow neither positive nor negative values to be privileged in algebraic equations. In other words, every equation which has a positive solution also has an algebraically indistinguishable negative solution (the complex conjugate). Thus, all our charges (but not necessarily masses) must exist in both positive and negative states. This is the precise requirement for the existence of antiparticles; even those particles, such as the neutron and neutrino, which have no electric charge still have antiparticles because they have strong and/or weak charges whose signs may be changed (under the process of charge conjugation, already mentioned).

4 Divisible and indivisible: space and charge versus time and mass

Now space is like time in being nonconserved, like mass in being real, and, apparently, like charge in being dimensional. Dimensionality, however, doesn't really look like a basic property. Is there any basic property which explains it? It seems to me very probable that there is, and, in looking at this question, it will be necessary once more to examine the relationship between space and time.

Space and time have often been assumed to be alike in most respects, but there is good evidence that they are fundamentally different. Space, for example, is always used in direct measurement; in fact, it is impossible to measure anything but space. Our 'time'-measuring devices, such as pendulums, mechanical clocks, and crystal and atomic oscillators, all use some concept of repetition of a spatial interval. Special conditions have to be used to set up such measurements, whereas any object whatsoever can be used to measure space. Space also is reversible – and it is this reversibility which is used in the measurement of time – but time is not.

Perhaps, we might find it convenient here to go back a couple of millennia and look at the famous paradoxes of Zeno of Elea. Of course, many people think these have been answered by the use of limits or infinite series, but Whitrow, who has made the most extensive and influential recent study, thinks otherwise.[2] In the well-known argument about the race between Achilles and the Tortoise, Achilles, in any number of time intervals, should never catch up with the Tortoise, to whom he has given a lead, because, each time he thinks he has caught up, he finds the Tortoise has already moved further ahead, even if only by an ever smaller amount. Another example is the Dichotomy Paradox, in which an object moving over any distance can never get started because it must cover half the distance before it covers the whole, and a quarter of the distance before it covers half, and so on; to go any distance in a finite amount of time, it must already have been involved in an infinite number of operations.

The problem seems to be the infinite divisibility of time; Achilles, for example, never catches the Tortoise because we have assumed that the time for the race can be divided up into finite intervals. On the basis of these, and similar paradoxes, Whitrow writes: 'One can, therefore, conclude that the idea of the infinite divisibility of time must be rejected, or ... one must recognize that it is ... a logical fiction.'[3] And the more recent authors, Peter Coveney and Roger Highfield conclude that: 'Either one can seek to deny the notion of 'becoming', in which case time assumes essentially space-like properties; or one must reject the asumption that time, like space, is infinitely divisible into ever smaller portions.'[4] The paradoxes seem to show, according to Whitrow, that motion is 'impossible if time (and, correlatively, space) is divisible ad infinitum'.[5]

Zeno's paradoxes are not just sophistry; Bertrand Russell considered them 'immeasurably subtle and profound', and A. N. Whitehead thought that they showed an 'instant of time' to be 'nonsense'.[6] Our reason for including them here is to show that there is good evidence that one cannot simply assume that time can be indefinitely subdivided like space. There is every reason to believe, in fact, that time, unlike space, is an absolute continuum. There is no infinite succession of measurable instants in time, as supposed in the paradoxes, because there are no instants. Time cannot actually be divided. To use a more contemporary jargon, space is digital, time is analogue – and we have both concepts in nature because we have both parameters.[7]

Continuity is a word with many meanings, and different uses of the word have caused confusion. The 'continuity' attributed to space because of its indefinite divisibility is not what is meant by the absolute continuity of time. Absolute continuity cannot be visualised and any process used to describe it would deny continuity. The property which space has that is often referred to as 'continuity' is indefinite elasticity, its 'continual' recountability or its unending divisibility. But the very divisibility of space is what denies it *absolute* continuity; and the elastic nature of the divisibility comes from the entirely different property of nonconservation. We expect a nonconserved quantity to have nonfixed units, but they are units nonetheless. The whole process of measurement depends crucially on the divisibility of space, or creation of discontinuities within it. Thus the entire problem of Zeno's paradoxes disappears as soon as we accept that we can have discontinuities or divisibility in space, but not in time.

Space can be discontinuous in both quantity and direction; it can be reversed and changed in orientation; and, without both of these properties, measurement would be impossible. Time, however, cannot be reversed, precisely because it is absolutely continuous. Any reversal of time would require discontinuity. For the same reason, time cannot be multidimensional, or, in our terminology, 'dimensional'. The same distinction occurs between mass and charge. Mass is an absolute continuum present in all systems and at every point in space (if only in the form of fields and energy); this is why there is no negative mass, for negative mass would necessarily require a break in the continuum. Charge, on the other hand, is divisible and observed in units; of course, because charge is a conserved quantity, unlike space, these units must be fixed, unlike those of space. Again, charge as a noncontinuous quantity is also dimensional, and, thus we might suggest divisibility as the 'cause' of dimensionality. Though we cannnot easily prove that divisibility causes dimensionality, we can at least see why absolutely continuous quantities are *non*dimensional.

One often reads, of course, about a 'reversibility paradox', where time, according to the laws of physics is reversible in mathematical sign, when it is clearly not reversible in physical consequences. Time, however, we need to remember, is characterised by imaginary numbers, and imaginary numbers are not privileged according to sign. Thus, it is quite possible to have a time which has equal positive and negative mathematical solutions because it is imaginary, but which has only one physical direction because it is continuous. (The corresponding unipolarity, or single sign, of mass is the reason why we have a CPT, rather than an MCPT, theorem, C standing for charge conjugation, P for space reflection and T for time reversal, all of which have two mathematical sign options.)

The distinction between space and time has many interesting consequences. In principle, when we mathematically combine space and time in Minkowski's 4-vector, as symmetry apparently requires us to do, we have two options: we can either make time space-like (or discrete) or space time-like (or continuous). This seems to be the origin of wave-particle duality. The discrete options lead to particles, special relativity and Heisenberg's quantum mechanics. The continuous options lead to waves, Lorentzian relativity and Schrödinger's wave mechanics. Heisenberg makes everything discrete, so mass becomes charge-like quanta in quantum mechanics; Schrödinger, on the other hand, makes everything continuous, so charge becomes mass-like wavefunctions in wave mechanics. In measurement, the true situations are restored, for Heisenberg reintroduces continuous mass via the uncertainty principle and the virtual vacuum, while Schrödinger reintroduces discrete charge via the collapse of the wavefunction.[8]

Another aspect of the distinction between space and time occurs in the fundamental fact that time, in the definition of velocity and acceleration, the basic quantities used in dynamics, is the independent variable, whereas space is the dependent variable. This situation arises because time, unlike space, is not susceptible to measurement. We have no control over the variation of time, and so its variation is necessarily independent.

The fundamental distinction between the status of space and time almost certainly also has relevance in mathematics. In the seventeenth century, there were two processes of differentiation: the discrete (or Leibnizian), essentially modelled on variation in space; and the continuous (or Newtonian), essentially modelled on variation in time. Like particles and waves, each is a valid option, for differentiation is a property linked to nonconservation, and not concerned, in principle, with the difference between absolute continuity and indefinite divisibility. (The solutions of Zeno's paradoxes that invoke the concept of limit tacitly assume the Newtonian definition of differentiation.) Again, it is probable that the Cantorian definition of an absolutely continuous set of real numbers has equal validity with the idea of an infinitely constructible, though not absolutely continuous, set of real numbers based on algorithmic processes.[9] The mathematical options that are available, here and elsewhere, are almost certainly a reflection of the availability of physical options. Continuity and discontinuity, finiteness and infinity, and so on, probably exist as mathematical categories because they are also physical categories.

5 A group of order 4

From what we have seen, then, the four basic parameters seem to be distributed between three sets of opposing paired categories: real / imaginary, conserved / nonconserved, divisible / indivisible, with each parameter paired off with a different partner in each of the categories:

space	real	nonconserved	divisible
time	imaginary	nonconserved	indivisible
mass	real	conserved	indivisible
charge	imaginary	conserved	divisible

The properties where they match, seem to be exactly identical, and where they oppose, to be in exact opposition. (Certain representations, however, like the Dirac equation involve mathematical reversals of physical properties, the Lorentz-invariant structure demanding either timelike space or spacelike time, with corresponding reversals in the properties of mass or charge.) Mathematically, this scheme incorporates a group of order 4, in which any parameter can be the identity element and each is its own inverse.[10]

An algebraic representation is easily accomplished by representing the properties of space (real, nonconserved, divisible) by, say, a, b, c, with the opposing properties (imaginary, conserved, indivisible) represented by $-a$, $-b$, $-c$. The group now becomes:

space	a	b	c
time	$-a$	b	$-c$
mass	a	$-b$	$-c$
charge	$-a$	$-b$	c

With group multiplication rules of the form:

$$a * a = -a * -a = a$$
$$a * -a = -a * a = -a$$
$$a * b = a * -b = 0$$

and similarly for b and c, we can establish a group multiplication table of the form:

*	space	time	mass	charge
space	space	time	mass	charge
time	time	space	charge	mass
mass	mass	charge	space	time
charge	charge	mass	time	space

This is the characteristic multiplication table of the Klein-4 group, with space as the identity element and each element its own inverse. However, there is no reason to privilege space with respect to the other parameters, since the symbols a and $-a$, b and $-b$, c and $-c$ are arbitrarily selected, and any of the other three parameters may be made the identity by defining its properties as *a*, *b*, *c*. For example, if mass is made the identity element, then the group properties may be represented by:

space	a	$-b$	$-c$
time	$-a$	$-b$	c
mass	a	b	c
charge	$-a$	b	$-c$

and the multiplication table becomes:

*	mass	charge	time	space
mass	mass	charge	time	space
time	time	mass	space	time
charge	charge	space	mass	charge
space	space	time	charge	mass

Various further representations are possible, and seem to be relevant, in particular, to the mathematical structure of the Dirac equation. For example, the identity element, say mass, could be represented by the scalar part of a quaternion (1) and the other three terms by the imaginary operators ***i***, ***j***, ***k***, with the + and – values completing the (now cyclic) quaternion group structure.

*	1M	*i*C	*j*T	*k*S
1M	1M	*i*C	*j*T	*k*S
*i*C	*i*C	–1M	*k*S	–*j*T
*j*T	*j*T	–*k*S	–1M	–*i*C
*k*S	*k*S	*j*T	*i*C	–1M

It is important to recognise here that the quaternion operators are extrinsically derived and not an integral component of the parameters space, time, mass and charge. Though the addition of these operators creates a new group structure, this structure is a relation between new mathematical constructs and not between the parameters themselves; it also presupposes the validity of the original symmetry between the parameters

If the 3-dimensionality of charge and space is directly involved, the overall structure would require a quaternion and a 4-vector within another overall quaternion-type arrangement. This could be accomplished using an octonian, with sixteen members ($\pm 1m$, $\pm is$, $\pm je$, $\pm kw$, $\pm et$, $\pm fx$, $\pm gy$, $\pm hz$), though this is no longer a group. The nonassociativity of the dimensional terms in the octonian extension seems to be lost within terms which effectively cancel each other out, and are of no physical significance.

If charge is taken as the identity element, and is represented by a scalar, the remaining structure for time, space and mass (and, implicitly, the energy, momentum and mass operators) becomes that of the Dirac algebra and SU(5).[11] Such representations do not determine the properties of space, time, mass and charge. They exist because the group has four components, and can, therefore, be represented by a 4-component structure like quaternions, in which the link between elements is made by a binary operation (squaring); but the link between a group with four components and a 4-dimensional space-time or mass-charge may be in itself significant.

Using the postulated group as a working hypothesis, it becomes possible to explore possible constraints on the laws of physics, as a result of group properties (as is shown below). Another area to be investigated might be the way in which the relationship between the quaternion representation and the requirement of separate conservation for charges might affect the fundamental particle structures that are possible.[12]

6 Scaling relations

The group elements are required to be their own inverses, and to be each identities. In addition, the group multiplication rule (when all possible arrangements are taken into consideration) requires:

$$\text{charge* time} = \text{space* mass} .$$

A binary operation which makes this possible is the squared multiplication of units, such as already exists for space and time in the 4-vector combination and for mass and charge when they are combined in a quaternion. It is also inherent in the description of charge, time, space and mass as, respectively, quaternion (or possibly pseudovector), pseudoscalar, vector and scalar, that the units of their squared quantities must be comparable numerically. To create the necessary number of independent fundamental relationships, we need to define three scaling constants (or rather scaling parameters, since they need not be actually constant if they are known to vary according to some fixed rule). And since the system has inherent duality in making each quantity its own inverse, then we must define a relation between each quantity and the inverse of every other, for which one further scaling constant (or parameter) will suffice.[13]

The group relationship predicts that such fundamental constants must exist, while effectively ensuring that their individual values have no independent meaning. To relate these to familiar scales of measurement, we create them from combinations of the four historically-generated fundamental constants G, c, h (or $h / 2\pi$),

$4\pi\varepsilon_0$. (Here, for convenience, we assume that 'charge' has the electromagnetic value, though this is not a necessary assumption, and a grand unified value could be used instead; the actual 'values' of the constants are not particularly significant – only the fact that some such scaling must exist.)

We can now express the scaling relations between the units of space (r), time (t), mass (m), and charge (q) as follows (with the equality sign being interpreted as meaning 'equivalence'):

$$r = ict \qquad (1)$$

$$r = \frac{G}{c^2}\, m \qquad (2)$$

$$iq = (4\pi\varepsilon_0 G)^{1/2}\, m \qquad (3)$$

The respective imaginary and quaternion operators required by t and q are significant in determining the signs of their squared units. These operators are normally subsumed within the symbols t and q, but here they are added for emphasis.

The further relations between any parameter and the inverse of any other can all be derived from:

$$m = \frac{h}{c^2}\,\frac{1}{it} \qquad (4)$$

This last result is the one that we recognise as being responsible for quantization of energy and other physical properties. Quantization could thus be said to be a result of the fact that each parameter is its own inverse. Quantization and duality of scale are aspects of the same phenomenon.

The four independent scaling constants in the above scheme become c, (G / c^2), $(4\pi\varepsilon_0 G)^{1/2}$, and (h / c^2). These are merely the scaling relations between the units of each quantity, but the presence of c^2 and h informs us that these quantities are fundamental to physics, whether classical, relativistic or quantum. In principle, any term related to another by a scaling relation in a meaningful physical equation can be replaced by that term to produce another meaningful equation.

A significant aspect of the binary operation between parameters is the squaring of the units of each, or the multiplication of a unit of any parameter by an identically-valued unit of the same parameter. Now, units of mass and charge have individual identities, unlike those of space and time, and so the 'squaring' of their units becomes the multiplication of individual units, such as m_1m_2 and q_1q_2, and such 'squaring' must be a universal operation between any units of mass and charge, no individual unit being privileged. It will be convenient to give this process the name of 'interaction'.[14]

7 Constructed quantities

The most fundamental laws of physics are essentially definitions and conservation laws. Classical mechanics, for example, is structured on only two fundamental requirements: the construction of a quantity involving conserved and nonconserved parameters (force, energy, momentum, action, Lagrangian, Hamiltonian) and the definition of its behaviour under variation of the variable components, that is, whether it is defined to be zero, invariant, or an extremum. Essentially, the laws of classical mechanics are set up to define what is meant by a conservative system.

Now, the key concepts in classical mechanics (as in other aspects of physics) are those which combine the minimum information necessary to distinguish the conserved and nonconserved parameters. Of the conserved quantities, mass is universal and never zero, and therefore must be present; charge, however, is local, and can take zero values. To specify the conservation or invariability of mass, we also need to specify the nonconservation or variability of space and time; hence, these parameters are included in differential form. A convenient way to define a system, therefore, would be the construction of a quantity containing mass and the differentials of space and time. The most immediately useful constructs then include $p = m\, dr / dit$ and $F = dp / dit$ (with time, most conveniently specified as the independent variable). The second quantity, as has been previously explained, has the advantage of producing a real rather than an imaginary construct. Now, space, of course, is really a vector (neglecting, for the moment, any 4-vector aspects); to incorporate this aspect, we may multipy both terms by the unit vector $\mathbf{r} / r$, to yield the familiar quantities, *momentum*,

$$\mathbf{p} = m\,\frac{d\mathbf{r}}{dit}$$

and *force*,

$$\mathbf{F} = \frac{d\mathbf{p}}{dit}\,.$$

(Imaginary and quaternion labels are retained here for emphasis but would not, of course, normally be used.)

The definitions of such quantities are, as yet, purely mathematical and convey no additional physical information. This can now be supplied, however, by using the scaling relations to find other quantities to which these defined ones can be related, while at the same time applying the conditions for conservation and nonconservation. This enables us to set up a system of equations for classical mechanics and electromagnetic theory.

8 Classical mechanics

From (2) and (3), remembering that each element of mass is unique, we may derive the expression

$$Gm_1m_2 = h\frac{r}{it}$$

In differential form, under the specific conservation of mass elements,

$$Gm_1m_2 = h\frac{dr}{dit} ,$$

from which

$$\frac{Gm_1m_2}{c^2it} = m\frac{dr}{dit} = p .$$

The mass term on the right hand side, of course, is a new mass unit, distinguishable from m_1 and m_2.

By differentiation, and a further substitution,

$$-\frac{Gm_1m_2}{c^2\, i^2t^2} = -\frac{Gm_1m_2}{r^2} = \frac{dp}{dit} .$$

Applying the unit vector, **r** / *r*, this becomes

$$-\frac{Gm_1m_2}{r^3}\,\mathbf{r} = \frac{d\mathbf{p}}{dit} ,$$

which is a combination of Newton's law of gravitation and second law of motion, with the left hand side a new equivalent quantity for force, conventionally described as *gravitational force.*

Neither m_1 nor m_2 is, of course, privileged, and so the equation can also be written in the form:

$$-\frac{Gm_2m_1}{r^3}\,\mathbf{r} = \frac{d\mathbf{p}}{dit}$$

Interpreting the vectors **r** and **p** as directed from m_1 to m_2 means that reversing the mass terms produces reversed vectors, from m_2 to m_1, as required by Newton's third law of motion.

The equivalent case for charges defines Coulomb's law of electrostatics and introduces *electrostatic force* (with the opposite sign, and hence reversed vector, for identically valued charges):

$$\frac{q_1q_2}{4\pi\varepsilon_0 r^3}\,\mathbf{r} = \frac{d\mathbf{p}}{dit} .$$

All the other significant and relations of classical mechanics, in any of its forms, can be derived now by purely mathematical manipulation. For example, nterpreting a 'system' to mean any combination of unit masses, the conservation of momentum follows by integration of the total force over time, and the conservation of angular momentum ($\mathbf{L} = \mathbf{r} \times \mathbf{p}$) from the fact that $d\mathbf{p} / dt$ in a conservative system is zero.

Direct manipulation of the scaling relations reveals that momentum terms are equivalent to $mc\mathbf{r} / r$, and that scalar terms of the form Gm_1m_2 / r and $q_1q_2 / 4\pi\varepsilon_0 r$, which we may describe as *gravitational and electrostatic potential energies*, are equivalent to those of the form mc^2; in each of these cases, *m* may be described as an 'equivalent mass'. Though these results normally emerge only from relativity theory, they are actually inherent in the structure from which classical mechanics must be derived.

Further results follow from on immediately from the mathematical definition of new concepts. Thus, defining velocity as $\mathbf{v} = d\mathbf{r} / dit$ and acceleration as $\mathbf{a} = d\mathbf{v} / dit$, and field intensity as $\mathbf{F} / m$, we have, in the case of constant mass, $\mathbf{F} = m\mathbf{a}$, and can define *gravitational field intensity* as

$$\mathbf{g} = -\frac{Gm}{r^3}\,\mathbf{r}$$

and *electrostatic field intensity* as

$$\mathbf{E} = -\frac{iq}{4\pi\varepsilon_0 r^3}\,\mathbf{r} .$$

From vector theory, we can show that, for the related *scalar potentials*, $\phi = -Gm / r$ and $\phi = -iq / 4\pi\varepsilon_0 r$,

$$\mathbf{g} = -\nabla\phi$$

and

$$\mathbf{E} = -\nabla\phi ,$$

and, also, that the respective force laws are equivalent to the Laplace equations

$$-\nabla^2\phi = \nabla.\mathbf{g} = 0$$

and

$$-\nabla^2\phi = \nabla.\mathbf{E} = 0 ,$$

in a space without sources, and to the Poisson equations,

$$-\nabla^2\phi = \nabla.\mathbf{g} = 4\pi\rho G$$

and

$$-\nabla^2\phi = \nabla.\mathbf{E} = \rho / \varepsilon_0 ,$$

in a space with them. None of this requires any new physical argument.

9 Classical electromagnetic theory

To replace nonrelativistic equations with relativistic ones, we simply replace all vector terms with 4-vectors, **r**, for example, being replaced by (**r**, *ict*). This procedure can be done, of course, with purely mechanical equations, to generate the standard results of special relativity, but it is particularly significant in classical electromagnetic theory, which follows on immediately from applying 4-vector terms to the definition of electrostatic force.

The significant fact here is that charge is *locally* conserved, and, hence, by a standard argument, the continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla . \mathbf{j} = 0 ,$$

must apply, with ρ defined as the charge density and **j** = ρ **v** as the current density. The differential operator in this equation is a 4-vector, and so, recognisably, is the quantity with scalar and vector parts, ρ and **j** / *c*.

Now, the scalar part of this latter quantity appears in Poisson's equation,

$$- \nabla^2 \phi = \rho / \varepsilon_0 ,$$

which is the differential form of Coulomb's law, and so we should expect to find an equivalent vector part ($\mathbf{A} / c$) for φ, and an equivalent scalar part ($-(1/c^2)\, \partial^2 / \partial t^2$) for ∇^2. Since the new differential operator $\Box = ((1/c^2)\, \partial^2 / \partial t^2 - \nabla^2)$ is itself a universal scalar, we may separate out the scalar and vector parts of the total equation to give *the wave equations*:

$$\Box \phi = \rho / \varepsilon_0$$

and

$$\Box \mathbf{A} = \mathbf{j} / \varepsilon_0$$

It is significant that φ and **A** are arbitrary to the point where they satisfy these equations (the condition of gauge invariance, as previously discussed under nonconservation). For convenience, we can arbitrarily restrict the values using a *gauge condition*. If we choose the so-called Lorentz gauge, in which

$$\frac{\partial \phi}{\partial t} + \nabla . \mathbf{A} = 0 ,$$

and define new vectors **E** and **B**, without reference to physical characteristics, such that

$$\mathbf{E} = - \nabla \phi + \frac{\partial \mathbf{A}}{\partial t}$$

and

$$\mathbf{B} = \nabla \times \mathbf{A} ,$$

we obtain the four Maxwell equations in their standard form, and identify **E** as the electrostatic field vector.[15]

10 Conservation laws and fundamental symmetries

The outline derivations of classical mechanics and electromagnetic theory show that the group structure of space, time, mass and charge has the power to derive conventional results in a relatively simplified form. Numerous new mathematical results can be generated by even more direct uses of the symmetries. As previously noted, Noether's theorem requires the translation symmetry of time to be linked to the conservation of energy. Of course, since energy is related to mass by the equation $E = mc^2$, then the translation symmetry of time is also linked to the conservation of mass. To put it another way, the nonconservation of time is responsible for the conservation of mass. This is a result we could have derived from symmetry alone; and so, extending the analogy, we could link the conservation of the quantity of charge with the nonconservation, or translation symmetry of space; and since the latter is already linked with the conservation of linear momentum, we could propose a theorem in which the conservation of linear momentum was responsible for the conservation of the quantity of charge (of any type). By the same kind of reasoning, we could make the conservation of *type* of charge linked to the rotation symmetry of space, and so to the conservation of angular momentum, as in the following scheme:

symmetry	**conserved quantity**	**linked conservation**
space translation	linear momentum	value of charge
time translation	energy	value of mass
space rotation	angular momentum	type of charge

In fact, we can already show these principles to be true in special cases. As Fritz London showed in 1927, the conservation of electric charge within a system is identical to invariance under transformations of the electrostatic potential by a constant representing changes of phase, and the phase changes are of the kind involved in the conservation of linear momentum. Since, in a conservative system, electrostatic potential varies only with the spatial coordinates, this is, in effect, a statement of the principle that the quantity of electric charge is conserved because the spatial coordinates are not, which is a special case of the first predicted relation.

In the second case, there is the relation between spin and statistics observed in fundamental particles. Fermions and bosons have different values of spin angular momentum; and they also differ in that fermions probably carry weak units of charge, where bosons have none. It thus appears to be the presence of a particular *type* of charge which determines the angular momentum state of the particle, so conservation of this type of charge is linked with the value of angular momentum.

11 The Dirac equation

Yet another significant mathematical result follows from the basic representations of 4-vector space-time and quaternion mass-charge. A direct combination of these two constructs, putting the four parameters onto an equal overall footing in a single mathematical representation, produces a 32-part algebra which is identical in all respects to the 32-part algebra used in the Dirac equation for the electron, but much simpler in form and more powerful. (The 32 parts are made up of 2 complex numbers, 6 complex vectors, 6 complex quaternions and 18 complex vector quaternions (figure 3); terms equivalent to the five gamma matrices (for example, **i***i*, **j***i*, **k***i*, *j*, *i***k**) are easily derived.)[16]

2 complex numbers	$(1,i)$
6 complex unit vectors	$(1,i) \times (\mathbf{i},\mathbf{j},\mathbf{k})$
6 complex unit quaternions	$(1,i) \times (i,j,k)$
18 complex vector quaternions	$(1,i) \times (\mathbf{i},\mathbf{j},\mathbf{k}) \times (i,j,k)$

Figure 3. A 32-part algebra produced by combining 4-vectors and quaternions.

Once the Dirac algebra has been established, it is a relatively easy process to show that the Dirac equation follows from quantization of a basic classical conservation equation, and the algebra, in this case, becomes simplified to a virtually pure quaternion algebra, as the vector element is removed via a scalar product. We begin with the Lorentz-invariant relationship between energy, mass and momentum:

$$E^2 - p^2 - m^2 = 0 .$$

Factorization of this expression requires the use of a complex and noncommutative algebra, viz. quaternions:

$$(kE + iip + ijm)\,(kE + iip + ijm) = 0 .$$

We can also incorporate the factor $e^{-i(Et - \mathbf{p.r})}$ without requiring new physical information:

$$(kE + iip + ijm)\,(kE + iip + ijm)\, e^{-i(Et - \mathbf{p.r})} = 0$$

In the classical equation, E and p are variables within the requirement that $E^2 - p^2$ is a constant for fixed m^2. However, quantization changes the status of these terms so that, for stationary quantum states, E and p become fixed, along with m. The variability now becomes confined to the space and time parameters incorporated into the exponential term, which can now be seen, physically, to represent the entire group of space and time translations and rotations.

A more general variability of space and time can be incorporated by replacing the factor $(kE + iip + ijm)$ from the left with the differential operator

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)$$

acting on the 'wavefunction'

$$\psi = (kE + iip + ijm)\, e^{-i(Et - \mathbf{p.r})}$$

producing the expression

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)\psi = 0 ,$$

which we recognise as the Dirac equation (in a form already second quantized because the quantization process has been applied to both the differential operator and the wavefunction).[17]

It will be recognised that the need for quantization of E and p comes from the inverse relation between m and t, and that the process has a profound effect on the classical energy-conservation equation. In quantizing via the Dirac equation, and, at the same time, imposing Lorentz-invariance, we effectively restructure the properties of the physical quantities involved, though our *physical* interpretation of the quantities remains unchanged.[18]

In classical relativistic theory, we emphasize the 4-vector nature of $(iE, \mathbf{p})$ and describe $E^2 - p^2$ as an invariant, but, here we incorporate the invariance directly, and define a new term with five components, $(kE + iip + ijm)$, with quantized rest mass. This 5-'dimensional' quantity combines the effects of 3-dimensional conserved and nonconserved parameters (the momentum term p having 3 dimensions, although only one is normally defined). In effect, we structure mass (or energy-momentum-mass) as a 3-dimensional quantized and conserved parameter, like charge (with one of the 'dimensions' being itself dimensional). This is the result previously achieved by structuring charge-mass-space-time as a quaternion, with charge as the real or identity element.

12 Conclusion

The prediction of new mathematical theorems and the derivation of new algebraic concepts, as well as the procedures for obtaining standard theorems in classical mechanics, electromagnetic theory and quantum mechanics, show that the method of symmetry based on fundamental basic principles is not just a philosophical issue, but also a powerful method of generating new results, and of codifying existing ones. In fact, it is almost inevitable that new discoveries will follow after any successful exercise in getting down to the basics.

Notes and References

1 This is, of course, if we wish to retain associativity. One further extension exists in the 8-part octonians or Cayley numbers, which break associativity, and which are discussed later in the paper.
2 G. J. Whitrow, *The Natural Philosophy of Time* (Nelson, London, 1961), pp. 135-57
3 *op. cit.*, 152
4 P. Coveney and R. Highfield, *The Arrow of Time* (London, 1990), 28
5 *op. cit.*, 143
6 *ibid.*, 144, 157
7 We can also say that time is the set of reals with the standard topology superimposed (and is nonalgorithmic); space is the set of reals without the topology (and is algorithmic). Henri Bergson, according to Whitrow (*op. cit.*, 156-7), 'enthusiastically adopted the view' that time 'is wholly indivisible', 'as a means of escaping the difficulties raised by Zeno, concerning both temporal continuity and atomicity, without abandoning belief in the reality of time. ... Unfortunately, in attacking the geometrization (or spatialization) of time he went too far and argued that, because time is essentially different from space, therefore it is fundamentally irreducible to mathematical terms.'
8 With the altered parameters represented by S*, T*, M*, C*, the respective options are S, T*, M*, C (Heisenberg) and S*, T, M, C* (Schrödinger). Wave-particle duality is discussed in P. Rowlands, *Waves Versus Corpuscles: The Revolution That Never Was*, PD Publications, Liverpool, 1992; 'Quantum indeterminacy, wave-particle duality and the physical interpretation of relativity theory from first principles', *Proceedings of Conference on Physical Interpretations of Relativity Theory III*, British Society for Philosophy of Science, London, September 1992, 296-310; 'Quantum uncertainty, wave-particle duality and fundamental symmetries', in S. Jeffers, S. Roy, J-P. Vigier and G. Hunter (eds.), *The Present Status of the Quantum Theory of Light: A Symposium in Honour of Jean-Pierre Vigier* (*Fundamental Theories of Physics*, vol. 80, Kluwer Academic Publishers), 1997, 361-372.
9 We may note here the fundamental significance of the Löwenheim-Skolem theorem, that any consistent finite, formal theory has a denumerable model, with the elements of its domain in a one-to-one correspondence with the positive integers.
10 For related work, see P. Rowlands, 'The Fundamental Parameters of Physics', *Speculat. Sci. Tech.*, 6, 69-80, 1983; 'A new formal structure for deriving a physical interpretation of relativity', *Proceedings of Conference on Physical Interpretations of Relativity Theory II*, British Society for Philosophy of Science, London, September 1990, 264-8; *The Fundamental Parameters of Physics: An Approach towards a Unified Theory*, PD Publications, Liverpool, 1991, pp. 1-19.
11 P. Rowlands and J. P. Cullerne, 'A Symmetry Principle for Deriving Particle Structures', *Proceedings of ANPA XX*, Cambridge, September 1998.
12 *ibid.* See also P. Rowlands (1991), ref. 10, pp. 55-87.
13 The existence of the binary operation of squaring within the parameter group seems to be linked to the same operation being responsible for the 4-dimensionality of space-time and mass-charge.
14 It will be recognised that 'interaction' in this sense is nonlocal.
15 P. Rowlands (1991) (ref. 10), pp. 20-34.
16 A group version, with + and – units requires 64 terms (as does (S*, T*, M*, C*) × (S, T, M, C)).
17 P. Rowlands, 'An algebra combining vectors and quaternions: A comment on James D. Edmonds' paper', *Speculat. Sci. Tech.*, 17, 279-282, 1994; 'Some interpretations of the Dirac algebra', *Speculat. Sci. Tech.*, 19, 243-51, 1996; 'A New Algebra for Relativistic Quantum Mechanics', *Proceedings of Conference on Physical Interpretations of Relativity Theory V*, British Society for Philosophy of Science, London, September 1996, 381-7; 'The physical consequences of a new version of the Dirac equation', in G. Hunter, S. Jeffers and J.-P. Vigier (eds.), *Causality and Locality in Modern Physics and Astronomy: Open Questions and Possible Solutions*, Kluwer, 1998, 397-402; 'Further Considerations of the Dirac Algebra', *Proceedings of Conference on Physical Interpretations of Relativity Theory VI*, British Society for Philosophy of Science, London, September 1998.
18 See the works in refs. 8 and 11. In incorporating both the explicit quantization of *E-p-m* and the quaternion operators, the Dirac equation combines S* and T with an effective restructuring of M with the properties of C. Significantly, this has five components. The fact that only one direction of spin is well-defined is a consequence of using S* for S.

Mach's Principle

D. F. Roscoe, School of Mathematics,
Sheffield University, Sheffield, S3 7RH, UK.
Email: D.Roscoe@ac.shef.uk
Tel: 0114-2223791, Fax: 0114-2223739

April 23, 1999

Abstract

This article shows how, with minimal assumptions, it is possible to project a quasi-classical physical space and time out of a universe of primitive particles. In the limit of very large distances, this model space and time can be considered to equate with a Machian-type cosmology within which the inertial properties of the space and time structure are 'projected' from the material content of the universe. We present this structure as the appropriate physical background within which theories of gravitation in the regions of disturbing sources should be constructed.

1 Introduction

Although most reading this article will have a general understanding of Mach's Principle, its centrality to our argument makes a short review a worthwhile investment.

1.1 Mach's Principle: conventional approach

Briefly, there are two kinds of mass: gravitational mass and inertial mass. Gravitational mass is what is measured on any kind of weighing machine (classically a pan-balance, in which the mass-to-be-measured is weighed against a collection of standard masses) whilst inertial mass is what is measured in a collision experiment between the mass-to-measured and a standard mass. In each case, the measured quantity is measured *relative* to some chosen standard, and therefore has no absolute significance.

The relevant facts about inertia mass are best explained firstly in the context of collisions between two smooth balls on a horizontal smooth surface viewed from a non-accelerating frame of reference (the precise meaning of the adjectival verb 'non-accelerating' is given shortly): Suppose we arrange for two balls, A and B say, to be rolled along the same line at different speeds so that they collide, and then rebound (necessarily on the same line also), and that the *change* in the speeds of each of the balls is measured to be ΔV_A and ΔV_B respectively. Then it is found that the ratio $\Delta V_A / \Delta V_B$ always has the same value *independently* of the initial speeds of the two balls. In other words, the calculated ratio appears to be a relative property of the balls, rather than being dependent on the initial conditions of the experiment.

Now suppose that the experiment is repeated, but is now viewed from an accelerating frame of reference. It is now found that the ratio $\Delta V_A / \Delta V_B$ varies according to the initial speeds of the two balls.

The above paragraphs are clear, except in one respect: that is, the notions of 'accelerating/non-accelerating' are undefined. This lack of rigour is usually rectified by *defining* the state of non-acceleration to be relative to the distant galaxies: specifically, by identifying that frame of reference which appears to be at rest with respect to the statistically averaged motion of distant galaxies, and then using this very special frame as a standard against which non-accelerating motion is defined. It is then found that the ratio $\Delta V_A / \Delta V_B$ calculated in the collision experiments is constant in this class of frames; this ratio is termed *the relative inertial mass of the two balls*, and the frames within which it can be measured (the non-accelerating frames) are termed *inertial frames.*

The above analysis makes it clear that there is some kind of relationship between the distant galaxies, and the idea of relative inertial mass - and the statement of the existence of such a relationship is termed 'Mach's Principle'.

1.2 Difficulties arising from this conventional approach

The approach to Mach's Principle outlined above gives a very clear impression that locally defined inertial frames, and the property of inertial mass, are somehow connected to the distant masses of the universe. This has given rise to the general idea that, somehow, the collection of all material in the universe *induces* local inertial properties. Typically, then, mechanisms proposed to underly Mach's Principle conform to the following framework:

- Assume some notion of metrical space and time. This can be the space and time of our common senses (which is appropriate for the application of Newtonian mechanics and gravitation, for example), or it can be the spacetime of modern relativity theory;

- Postulate some relational mechanism which is assumed to exist between all material and which has the effect of inducing the property of 'inertial mass' in any given object when its effects are integrated over the whole universe. In a Newtonian framework, such mechanisms are assumed to be IAAD mechanisms, whilst in the 'modern' framework, they are assumed to be retarded field mechanisms.

In the present author's view, all such mechanisms are flawed in the same sense: which is that, by their assumed pre-existence of some form of physical space within which the proposed theory of Machian inertial induction is constructed, the fundamental nature of Mach's Principle is subverted. To understand why this is the case, we reconsider the concept of 'inertial frame'.

1.3 Inertial frames reconsidered

In the conventional approach, inertial frames are effectively defined with respect to the, so-called, universal rest-frame (which, incidently, can nowadays be defined with respect to the 2.7^o cosmic background). However, there is no need to proceed in this conventional way, since it is perfectly possible to give a totally laboratory-bound definition of the 'inertial frame': specifically, an inertial frame can be *defined* to be any frame of reference within which the series of collision experiments discussed above yields the ratio $\Delta V_A / \Delta V_B$ to be a constant independently of the experiment's initial conditions. If this constant ratio is then termed as the 'relative inertial mass of the two balls', then the whole idea of the inertial frame and inertial mass is arrived at without any reference whatsoever to 'distant galaxies'.

More significantly, this approach brings into the foreground *the* crucial point about Mach's Principle: which is that it is *impossible* to define inertial frames in the absence of material. It is this point which invalidates all those interpretations of Mach's Principle which follow the pattern of §1.2. By the same token, this author considers General Relativity, which allows an internally consistent discussion of empty 'inertial spacetime' - that of special relativity - to be similarly invalidated.

Before considering how to progress with our alternative view of the fundamental significance of Mach's Principle, we need to understand the root of the connection between inertia and gravitation, as it is expressed in the Weak Equivalence Principle.

1.4 The Weak Equivalence Principle

Suppose now that the two balls used in the collision experiment for the determination of their relative inertial mass are weighed separately on a weighing machine; then it is found that the ratio of their *weights* appears to be identical to their relative inertial mass. It is this comparison which gives rise to the, so-called, Weak Equivalence Principle, and to the understanding that inertia and gravitation are intimately connected in some way.

2 A qualitative description of the new approach

We have argued that the fundamental significance of Mach's Principle is that it is impossible to define inertial frames in the absence of material; or, as a generalization, we can say that it is *impossible* to conceive physical space and time in the absence of material. It follows from this that, if we are to arrive at a consistent and fundamental implementation of Mach's Principle, then we need a theory of the world according to which (roughly speaking) notions of space & time are somehow projected out of primary relationships between objects. In other words, notions of space & time are actually metaphors for these primary relationships. Our starting point is to consider the calibration of a radial measure which conforms to these ideas.

Consider the following perfectly conventional procedure which assumes that we 'know' what is meant by a given radial displacement, R say. On a large enough scale ($> 10^8$ light years), we can reasonably assume it is possible to write down a relationship describing the amount of mass contained within a given spherical volume: say

$$M = U(R), \tag{1}$$

where U is, in principle, determinable. Of course, a classical description of this type ignores the discrete nature of real material; however, overlooking this point, such a description is completely conventional and unremarkable. Because M obviously increases as R increases, then U is said to be monotonic, with the consequence that the above relationship can be inverted to give

$$R = G(M) \tag{2}$$

which, because (1) is unremarkable, is also unremarkable.

In the conventional view, (1) is logically prior to (2); however, it is perfectly possible to reverse the logical priority of (1) and (2) so that, in effect, we can choose to *define* the radial measure in terms of (2) rather than assume that it is known by some independent means. If this is done then, immediately, we have made it impossible to conceive of radial measure in the absence of material. With this as a starting point, we are able to construct a completely Machian Cosmology in a way outlined in the following sections.

3 A discrete model universe

We begin by constructing a model universe which has matter distribution and velocity distribution isotropy properties defined at a single origin, and then analyse the model universe from the point of view of a material observer situated at this origin. The conclusion is finally reached that motions on the large scale are necessarily purely inertial and arbitrarily directed from the point of view of the material observer. It then follows that, since the large scale motions have no kinematic or dynamic centre, the same description will be obtained by any other material observer who is *stationary* with respect to the first - a non-stationary state for the second observer would mean he would observe a dipole anisotropy in the velocity distribution which would invalidate the mode of analysis. The isotropy properties initially assumed at a single origin can therefore be considered to have a cosmological nature.

The model universe is defined so that

- *it consists of an infinity of identical, but labelled, discrete material particles which are primitive, and possessing no other properties beyond being material;*
- *'time' is to be understood, in a qualitative way, as a measure of process or ordered change' in the model universe;*
- *there is an origin about which the distributions of material particles and of material particle velocities are statistically isotropic - meaning that the results of sampling along some line of sight over some characteristic time are independent of the direction of line of sight;*
- *the distribution of material is statistically stationary - meaning that the results of sampling along an arbitrarily chosen line of sight over some characteristic time are independent of sampling epoch.*

Although concepts of invariant spatio-temporal measurement are implicitly assumed to exist in this model universe, we make no apriori assumptions about their precise nature, but require that meanings for these concepts should arise naturally from the structure of the model universe and from the following analysis.

3.1 The invariant calibration of a radial coordinate in terms of counting primitive objects.

At (2), we have already introduced, in a qualitative way, the idea that radial displacement can be *defined* in terms of the amount of material contained within a given sphere and, in this section, we seek to make this idea more rigorous. To this end, we note that the most primitive invariant that can be conceived is that based on the counting of objects in a countable set; we show how this fundamental idea can be used to define the concept of invariant distance in the model universe.

The isotropy properties assumed for the model universe imply that it is statistically spherically symmetric about the chosen origin. If, for the sake of simplicity, it is assumed that the characteristic sampling times over which the assumed statistical isotropies become exact are infinitesimal, then the idea of statistical spherical symmetry gives way to the idea of exact spherical symmetry - thereby allowing the idea of some kind of rotationally invariant radial coordinate to exist. As a first step towards defining such an idea, suppose only that the means exists to define a succession of nested spheres, $S_1 \subset S_2 \subset ... \subset S_p$, about the chosen origin; since the model universe with infinitesimal characteristic sampling times is stationary, then the flux of particles across the spheres is such that these spheres will always contain fixed numbers of particles, say $N_1, N_2, ..., N_p$ respectively.

Since the only invariant quantity associated with any given sphere, S say, is the *number* of massive particles contained within it, N say, then the only way to associate an invariant radial coordinate, r say, with S is to *define* it according to $r = r_0 f(N)$ where r_0 is a fixed scale-constant having units of 'length', and the function f is restricted by the requirements $f(N_a) > f(N_b)$ whenever $N_a > N_b$, $f(N) > 0$ for all $N > 0$, and $f(0) = 0$. To summarize, an invariant calibration of a radial coordinate in the model universe is given by $r = r_0 f(N)$ where:

- $f(N_a) > f(N_b)$ whenever $N_a > N_b$;
- $f(N) > 0$ for all $N > 0$ and $f(0) = 0$.

Once a radial coordinate has been invariantly calibrated, it is a matter of routine to define a rectangular coordinate system based upon this radial calibration; this is taken as done for the remainder of this paper.

3.2 The mass model

At this stage, since no notion of 'inertial frame' has been introduced then the idea of 'inertial mass' cannot be defined. However, we have assumed the model universe to be composed of a countable infinity of labelled - but otherwise indistinguishable - material particles so that we can associate with each individual particle a property called 'mass' which quantifies the amount of material in the particle, and is represented by a scale-constant, m_0 say, having units of 'mass'.

The radial parameter about any point is *defined* by $r = r_0 f(N)$; since this function is constrained to be monotonic, then its inverse exists so that, by definition, $N = f^{-1}(r/r_0)$; suppose we now introduce the scale-constant m_0, then $Nm_0 = m_0 f^{-1}(r/r_0) \equiv M(r)$ can be *interpreted* as quantifying the total amount of material inside a sphere of radius r centred on the assumed origin. Although $r = r_0 f(N)$ and $M(r) = Nm_0$ are equivalent, the development which follows is based upon using $M(r)$ as a description of the mass distribution given as a function of an invariant distance parameter, r, of undefined calibration.

It is clear from the foregoing discussion that r is defined as a necessarily discrete parameter. However, to enable the use of familiar techniques, it will hereafter be supposed that r represents a continuum - it being understood that a fully consistent treatment will require the use of discrete mathematics throughout.

4 The absolute magnitudes of arbitrary displacements in the model universe

We have so far defined, in general terms, an invariant radial coordinate calibration procedure in terms of the radial distribution of material valid from the assumed origin, and have noted that such a procedure allows a routine definition of orthogonal coordinate axes. Whilst this process has provided a means by which arbitrary displacements can be described relative to the global material distribution, it does not provide the means by which an invariant *magnitude* can be assigned to such displacements - that is, there is no metric defined for the model universe. In order to understand how such a metric can be defined, we begin by noting the following empirical circumstances:

- An observer recognizes the fact of a spatial displacement by reference to his changed perspective of the universe of material;

- An observer judges the magnitude of a displacement in terms of the magnitude of the changes in the perspective of the material distribution.

These circumstances suggest that, if we wish to associate an absolute magnitude to a coordinate displacement in the context of the mass model, $M(r)$, then the initial requirement is to give a quantitative meaning to the notion of a mass-model 'perspective'. In general terms, a 'perspective' implies the existence of an observed object, and a particular angle-of-view onto the object. If, in the context of the mass-model, the observed object is considered defined by the specification of a constant-mass surface ($r = constant$) then, subject to the magnitude of the normal gradient vector, ∇M, being a monotonic function of r, total perspective information is precisely carried by the normal gradient vector itself. To see this, we note that the assumed monotonicity of the magnitude of ∇M means it is in a 1:1 relation with r; consequently this magnitude defines *which* constant-mass surface is observed. Simultaneously, the direction of ∇M, which is always radial, defines an angle-of-view onto this constant-mass surface. So, to summarize, an observer's perspective of the mass model, $M(r)$, can be considered defined by the normal gradient vector, $\mathbf{n} \equiv \nabla M$, at the observer's position.

We now consider the change in perspective arising from an infinitesimal change in coordinate position: defining the components of the normal gradient vector (the perspective) as $n_a \equiv \nabla_a M$, $a = 1, 2, 3$, then the *change* in perspective for a coordinate displacement $d\mathbf{r} \equiv (dx^1, dx^2, dx^3)$ is given by

$$dn_a = \nabla_j(\nabla_a M)dx^j \equiv g_{ja}dx^j, \quad g_{ab} \equiv \nabla_a \nabla_b M, \tag{3}$$

for which it is assumed that the geometrical connections required to give this latter expression an unambiguous meaning will be defined in due course. Given that g_{ab} is non-singular, we now note that (3) provides a 1:1 relationship between the contravariant vector dx^a (defining change in the observer's coordinate position) and the covariant vector dn_a (defining the corresponding change in the observer's perspective). It follows that we can define dn_a as the covariant form of dx^a, so that g_{ab} automatically becomes the mass model metric tensor. The scalar product $dS^2 \equiv dn_i dx^i$ is then the absolute magnitude of the coordinate displacement, dx^a, defined relative to the change in perspective arising from the coordinate displacement.

The units of dS^2 are easily seen to be those of *mass* only and so, in order to make them those of *length*^{2} - as dimensional consistency requires - we define the working invariant as $ds^2 \equiv (2r_0^2/m_0)dS^2$, where r_0 and m_0 are scaling constants for the distance and mass scales respectively and the numerical factor has been introduced for later convenience.

Finally, if we want

$$ds^2 \equiv \left(\frac{r_0^2}{2m_0}\right) dn_i dx^i \equiv \left(\frac{r_0^2}{2m_0}\right) g_{ij} dx^i dx^j \tag{4}$$

to behave sensibly in the sense that $ds^2 = 0$ only when $d\mathbf{r} = 0$, then we must replace the condition of non-singularity of g_{ab} by the condition that it is strictly positive (or negative) definite; in the physical context of the present problem, this will be considered to be a self-evident requirement.

4.1 The connection coefficients

We have assumed that the geometrical connection coefficients can be defined in some sensible way. To do this, we simply note that, in order to define conservation laws (ie to do physics) in a Riemannian space, it is necessary to be have a generalized form of Gausses' divergence theorem in the space. This is certainly possible when the connections are defined to be the metrical connections, but it is by no means clear that it is ever possible otherwise. Consequently, the connections are assumed to be metrical and so g_{ab}, given at (3), can be written explicitly as

$$g_{ab} \equiv \nabla_a \nabla_b M \equiv \frac{\partial^2 M}{\partial x^a \partial x^b} - \Gamma^k_{ab} \frac{\partial M}{\partial x^k}, \tag{5}$$

where Γ^k_{ab} are the Christoffel symbols, and given by

$$\Gamma^k_{ab} = \frac{1}{2} g^{kj} \left(\frac{\partial g_{bj}}{\partial x^a} + \frac{\partial g_{ja}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^j} \right).$$

5 The metric tensor given in terms of the mass model

It can be shown how, for an arbitrarily defined mass model, $M(r)$, (5) can be exactly resolved to give an explicit form for g_{ab} in terms of such a general $M(r)$: Defining

$$\mathbf{r} \equiv (x^1, x^2, x^3), \quad \Phi \equiv \frac{1}{2} < \mathbf{r}|\mathbf{r} > \quad \text{and} \quad M' \equiv \frac{dM}{d\Phi}$$

where $< \cdot|\cdot >$ denotes a scalar product, then it is found that

$$g_{ab} = A\delta_{ab} + Bx^a x^b, \tag{6}$$

where

$$A \equiv \frac{d_0 M + m_1}{\Phi}, \quad B \equiv -\frac{A}{2\Phi} + \frac{d_0 M' M'}{2A\Phi}.$$

for arbitrary constants d_0 and m_1 where, as inspection of the structure of these expressions for A and B shows, d_0 is dimensionless and m_1 has dimensions of mass. Noting that M always occurs in the form $d_0 M + m_1$, it is convenient to write $\mathcal{M} \equiv d_0 M + m_1$, and to write A and B as

$$A \equiv \frac{\mathcal{M}}{\Phi}, \quad B \equiv -\left(\frac{\mathcal{M}}{2\Phi^2} - \frac{\mathcal{M}'\mathcal{M}'}{2d_0\mathcal{M}}\right). \tag{7}$$

6 Geodesic distance determined in terms of the matter distribution

In this section, we show the remarkable result that the invariant physical radius of a sphere centred on the chosen origin in the model universe varies as the square root of the mass contained within the sphere.

Using (6) and (7) in (4), and after using $x^i dx^i \equiv r dr$ and $\Phi \equiv r^2/2$, we find, for an arbitrary displacement,

$$ds^2 = \left(\frac{r_0^2}{2m_0}\right)\left\{\frac{\mathcal{M}}{\Phi} dx^i dx^i - \Phi\left(\frac{\mathcal{M}}{\Phi^2} - \frac{\mathcal{M}'\mathcal{M}'}{d_0\mathcal{M}}\right) dr^2\right\}.$$

Now suppose that the displacement is purely radial; in this case, we find

$$ds^2 = \left(\frac{r_0^2}{2m_0}\right)\left\{\Phi\left(\frac{\mathcal{M}'\mathcal{M}'}{d_0\mathcal{M}}\right)dr^2\right\}.$$

Use of $\mathcal{M}' \equiv d\mathcal{M}/d\Phi$ reduces this latter relationship to

$$ds^2 = \frac{r_0^2}{d_0 m_0}\left(d\sqrt{\mathcal{M}}\right)^2 \quad \rightarrow \quad ds = \frac{r_0}{\sqrt{d_0 m_0}}d\sqrt{\mathcal{M}},$$

which defines the invariant magnitude of an infinitesimal *radial* displacement purely in terms of $\mathcal{M} \equiv d_0 M + m_1$, which represents the mass model. From this, we easily see that if we make the association $r \equiv s$ so that the radial coordinate r effectively coincides with the geodesic distance, then geodesic radial displacement from the chosen coordinate origin is *defined* by

$$r = \frac{r_0}{\sqrt{d_0 m_0}}\left(\sqrt{\mathcal{M}} - \sqrt{\mathcal{M}_0}\right),$$

where $\mathcal{M}_0$ is the value of $\mathcal{M}$ at $r = 0$; the significance of this result lies in the fact that it says the perception of physical displacement is *created* by the matter distribution.

For convenience, this result is restated as follows: since $M_0 = 0$ necessarily, then $\mathcal{M}_0 = m_1$ from which the above result can be equivalently arranged as

$$\mathcal{M} = \left[\frac{\sqrt{d_0 m_0}}{r_0}r + \sqrt{m_1}\right]^2. \tag{8}$$

Using $\mathcal{M} \equiv d_0 M + m_1$, then the mass-distribution function can be expressed in terms of the invariant radial displacement as

$$M = m_0\left(\frac{r}{r_0}\right)^2 + 2\sqrt{\frac{m_0 m_1}{d_0}}\left(\frac{r}{r_0}\right) \tag{9}$$

which, for large r, can be approximated as $M \approx m_0(r/r_0)^2$. It is clear from this latter relation that, on a large enough scale, the value m_0/r_0^2 is a global constant. For the remainder of this paper, the notation $g_0 \equiv m_0/r_0^2$ is employed.

7 A quasi-fractal mass distribution law, $M \approx r^2$: the evidence

A basic assumption of the *Standard Model* is that, on some scale, the universe is homogeneous; however, in early responses to suspicions that the accruing data was more consistent with Charlier's conceptions of an hierarchical universe (Charlier, 1908, 1922, 1924) than with the requirements of the *Standard Model*, de Vaucouleurs (1970) showed that, within wide limits, the available data satisfied a mass distribution law $M \approx r^{1.3}$, whilst Peebles (1980) found $M \approx r^{1.23}$. The situation, from the point of view of the *Standard Model*, has continued to deteriorate with the growth of the data-base to the point that, (Baryshev et al (1995))

> *...the scale of the largest inhomogeneities (*discovered to date*) is comparable with the extent of the surveys, so that the largest known structures are limited by the boundaries of the survey in which they are detected.*

For example, several recent redshift surveys, such as those performed by Huchra et al (1983), Giovanelli and Haynes (1986), De Lapparent et al (1988), Broadhurst et al (1990), Da Costa et al (1994) and Vettolani et al (1994) etc have discovered massive structures such as sheets, filaments, superclusters and voids, and show that large structures are common features of the observable universe; the most significant conclusion to be drawn from all of these surveys is that the scale of the largest inhomogeneities observed is comparable with the spatial extent of the surveys themselves. So, to date, evidence that the assumption of homogeneity in the universe is realistic does not exist. By contrast, and as the Baryshev review article shows, analyses of both pencil-beam and wide-angle surveys provide very strong support for the idea that the distribution of galaxies conforms to the power law $M \propto r^2$ up to the sample limits. More recent publications supporting this picture are those of Montuori & Labini (1997), Montuori, Labini & Amici (1997), Amendola et al (1997). As one might expect, since the fractal picture is, per se, contrary to the ideas of homogeneity with underly the big-bang picture, there are articles representing dissenting points of view. A typical one of these is that of Borgani (1995).

To summarize, there is considerable debate centered around the question of whether or not the material in the universe is distributed fractally or not, with supporters of the big-bang picture arguing that, basically, it is not, whilst the supporters of the fractal picture argue that it is with the weight of evidence supporting $M \propto r^2$. This latter position corresponds exactly with the picture predicted by the present approach.

8 The temporal dimension

Since, in its most general definition, time is a parameter which orders change within a system, then a necessary pre-requisite for its definition in the model universe is a notion of change within that universe, and the only kind of change which can be defined in such a simple place as the model universe is that of internal change arising from the spatial displacement of particles. Furthermore, since the system is populated solely by primitive baryonic particles which possess only the property of mass then, in effect, all change is gravitational change. This fact is incorporated into the cosmology to be derived by constraining all particle displacements to satisfy the Weak Equivalence Principle. We are then led to a Lagrangian description of particle motions in which the Lagrange density is degree zero in its temporal-ordering parameter. From this, it follows that the corresponding Euler-Lagrange equations form an *incomplete* set.

The origin of this problem traces back to the fact that, because the Lagrangian density is degree zero in the temporal ordering parameter, it is then invariant with respect to any transformation of this parameter which preserves the ordering. This implies that, in general, temporal ordering parameters cannot be identified directly with physical time - they merely share one essential characteristic. This situation is identical to that encountered in the Lagrangian formulation of General Relativity; there, the situation is resolved by defining the concept of 'particle proper time'. In the present case, this is not an option because the notion of particle proper time involves the prior definition of a system of observer's clocks - so that some notion of clock-time is factored into the prior assumptions upon which the theory is built.

In the present case, it turns out that the isotropies already imposed on the system conspire to provide an automatic resolution of the problem which is consistent with the already assumed interpretation of 'time' as a measure of ordered change in the model universe. To be specific, it turns out that the elapsed time associated with any given particle displacement is proportional, via a scalar field, to the invariant spatial measure attached to that displacement. Thus, physical time is defined directly in terms of the invariant measures of *process* with the model universe.

9 Dynamical constraints in the model universe

The model universe is populated exclusively by primitive baryonic particles which possess solely the property of mass. Consequently, all motions in the model universe are effectively gravitational, and we model this circumstance by constraining all such motions to satisfy the Weak Equivalence Principle by which we mean that the trajectory of a body is independent of its internal constitution.

Given that the distribution of particles is isotropic then, from the point of view of the chosen origin within which the velocity distribution of particles in the model universe also appears isotropic, symmetry arguments lead to the conclusion that this implies the net action of the whole universe of particles acting on any given single particle is such that the net acceleration of the particle must always *appear* to be directed through the coordinate origin. Note that this conclusion is *independent* of any notions of retarded or instantaneous action. These dynamical constraints can then be stated as:

> **C1** *Particle trajectories are independent of the specific mass values of the particles concerned;*
>
> **C2** *The acceleration of any particular massive particle is along the line connecting the particular particle to the coordinate origin.*

10 Gravitational trajectories

Suppose p and q are two arbitrarily chosen point coordinates on the trajectory of the chosen particle, and suppose that (4) is integrated between these points to give the scalar invariant

$$I(p,q) = \int_p^q \left(\frac{1}{\sqrt{2g_0}}\right) \sqrt{dn_i dx^i} \equiv \int_p^q \left(\frac{1}{\sqrt{2g_0}}\right) \sqrt{g_{ij} dx^i dx^j}. \qquad (10)$$

Then, in accordance with the foregoing interpretation, $I(p,q)$ gives a scalar record of how the particle has moved between p and q defined with respect to the particle's continually changing relationship with the mass model, $M(r)$.

Now suppose $I(p,q)$ is minimized with respect to choice of the trajectory connecting p and q, then this minimizing trajectory can be interpreted as a geodesic in the Riemannian space which has g_{ab} as its metric tensor. Given that g_{ab} is defined in terms of the mass model $M(r)$ - the existence of which is independent of any notion of 'inertial mass', then the existence of the metric space, and of geodesic curves within it, is likewise explicitly independent of any concept of inertial-mass. It follows that the identification of the particle trajectory $\mathbf{r}$ with these geodesics means that particle trajectories are similarly independent of any concept of inertial mass, and can be considered as the modelling step defining that general subclass of trajectories which conform to that characteristic phenomenology of gravitation defined by condition **C1** of §9.

11 The equations of motion

Whilst the mass distribution, represented by $\mathcal{M}$, has been explicitly determined in terms of the geodesic distance at (8), it is convenient to develop the theory in terms of unspecified $\mathcal{M}$.

The geodesic equations in the space with the metric tensor (6) can be obtained, in the usual way, by defining the Lagrangian density

$$\mathcal{L} \equiv \left(\frac{1}{\sqrt{2g_0}}\right) \sqrt{g_{ij}\dot{x}^i\dot{x}^j} = \left(\frac{1}{\sqrt{2g_0}}\right) \left(A < \dot{\mathbf{r}}|\dot{\mathbf{r}} > +B\dot{\Phi}^2\right)^{1/2}, \qquad (11)$$

where $\dot{x}^i \equiv dx^i/dt$, etc., and writing down the Euler-Lagrange equations

$$2A\ddot{\mathbf{r}}+\left(2A'\dot{\Phi}-2\frac{\dot{\mathcal{L}}}{\mathcal{L}}A\right)\dot{\mathbf{r}}+\left(B'\dot{\Phi}^2+2B\ddot{\Phi}-A'<\dot{\mathbf{r}}|\dot{\mathbf{r}}>-2\frac{\dot{\mathcal{L}}}{\mathcal{L}}B\dot{\Phi}\right)\mathbf{r}=0, \quad (12)$$

where $\dot{\mathbf{r}} \equiv d\mathbf{r}/dt$ and $A' \equiv dA/d\Phi$, etc. By identifying particle trajectories with geodesic curves, this equation is now interpreted as the equation of motion, referred to any centre of isotropic symmetry, of a single particle satisfying condition **C1** of §9.

However, noting that the variational principle, (10), is of order zero in its temporal ordering parameter, we can conclude that the principle is invariant wrt arbitrary transformations of this parameter; in turn, this means that the temporal ordering parameter cannot be identified with physical time. This problem manifests itself mathematically in the statement that the equations of motion (12) do not form a complete set.

12 The quantitative definition of physical time

12.1 Completion of equations of motion

Consider **C2**, which states that particle accelerations are directed through the coordinate origin. This latter condition simply means that the equations of motion must have the general structure

$$\ddot{\mathbf{r}} = G(t, \mathbf{r}, \dot{\mathbf{r}})\mathbf{r},$$

for scalar function $G(t, \mathbf{r}, \dot{\mathbf{r}})$. In other words, (12) satisfies condition **C2** if the coefficient of $\dot{\mathbf{r}}$ is zero, so that

$$\left(2A'\dot{\Phi}-2\frac{\dot{\mathcal{L}}}{\mathcal{L}}A\right)=0 \quad \rightarrow \frac{A'}{A}\dot{\Phi}=\frac{\dot{\mathcal{L}}}{\mathcal{L}} \quad \rightarrow \quad \mathcal{L}=k_0A, \quad (13)$$

for arbitrary constant k_0 which is necessarily positive since $A > 0$ and $\mathcal{L} > 0$. This latter condition, which guarantees (**C2**), can be considered as the condition required to close the incomplete set (12).

12.2 Physical time defined quantitatively as process

Equation (13) can be considered as that equation which removes the pre-existing arbitrariness in the 'time' parameter by *defining* physical time:- from (13) and (11) we have

$$\mathcal{L}^2 = k_0^2 A^2 \rightarrow A < \dot{\mathbf{r}}|\dot{\mathbf{r}} > + B\dot{\Phi}^2 = 2g_0 k_0^2 A^2 \rightarrow g_{ij}\dot{x}^i\dot{x}^j = 2g_0 k_0^2 A^2 \quad (14)$$

so that, in explicit terms, physical time is *defined* by the relation

$$dt^2 = \left(\frac{1}{2g_0 k_0^2 A^2}\right) g_{ij} dx^i dx^j. \quad (15)$$

In short, the elapsing of time is given a direct physical interpretation in terms of the process of *displacement* in the model universe.

13 The equations of motion: potential form

From §12, when (13) is used in (12) there results

$$2A\ddot{\mathbf{r}} + \left(B'\dot{\Phi}^2 + 2B\ddot{\Phi} - A' < \dot{\mathbf{r}}|\dot{\mathbf{r}} > -2\frac{A'}{A}B\dot{\Phi}^2\right)\mathbf{r} = 0. \quad (16)$$

If a potential function is defined as $V \equiv C_0 - < \dot{\mathbf{r}}|\dot{\mathbf{r}} > /2$, where C_0 is the arbitrary constant usually associated with the definition of a potential function, then, by (14)

$$V \equiv C_0 - \frac{1}{2} < \dot{\mathbf{r}}|\dot{\mathbf{r}} >= C_0 - \frac{v_0^2}{4d_0^2\, g_0}A + \frac{B}{2A}\dot{\Phi}^2, \quad (17)$$

where A and B are defined at (7). It can be shown how (16) can be expressed as

$$\ddot{\mathbf{r}} = -\frac{dV}{dr}\hat{\mathbf{r}}, \quad (18)$$

for unit vector, $\hat{\mathbf{r}}$.

Now notice that, since integrating the radial component of this latter gives directly $\dot{r}^2 + r^2\dot{\theta}^2 = -2V$, which must agree with (14), then the arbitrary constant, C_0, given in the definition of V at (17), must be identically *zero*; consequently, the potential function, V, is defined relative to an absolute background.

14 Discussion

We now consider the extent to which (18) reproduces Newtonian gravitational phenomena. Briefly, if we set $m_1 = 0$ in (8), then it is easily shown that $V = constant$, so that the universe is everywhere in dynamic equilibrium; a detailed analysis then shows that the case of *exact* Newtonian gravitation comes out of (18) as a one-point vanishingly small perturbation of this equilibrium universe. This equilibrium material universe (which is an exactly fractal $D = 2$ universe) can then be considered as the appropriate 'inertial background' within which to study the gravitational behaviour of variously shaped material distributions - for example, spiral galaxies.

References

[1] Amendola, L., D Nella, H., Montuori, M., Labini, F.S., 1997, Fractals - An Interdisciplinary Journal On The Complex Geometry Of Nature, 5, 635

[2] Baryshev, Yu V., Sylos Labini, F., Montuori, M., Pietronero, L., 1995, Vistas in Astronomy 38, 419

[3] Borgani, S., 1995, Physics Reports - Review Section Of Physics Letters, 251, 2

[4] Broadhurst, T.J., Ellis, R.S., Koo, D.C., Szalay, A.S., 1990, Nat 343, 726

[5] Broadhurst, T.J., Ellis, R.S., Glazebrook, K., 1990, Nat 355, 55

[6] Charlier, C.V.L., 1908, Astronomi och Fysik 4, 1

[7] Charlier, C.V.L., 1922, Ark. Mat. Astron. Physik 16, 1

[8] Charlier, C.V.L., 1924, PASP 37, 177

[9] Da Costa, L.N., Geller, M.J., Pellegrini, P.S., Latham, D.W., Fairall, A.P., Marzke, R.O., Willmer, C.N.A., Huchra, J.P., Calderon, J.H., Ramella, M., Kurtz, M.J., 1994, ApJ 424, L1

[10] De Lapparent, V., Geller,M.J., Huchra, J.P., 1988, ApJ 332, 44

[11] De Vaucouleurs, G., 1970, Sci 167, 1203

[12] Giovanelli, R., Haynes, M.P., Chincarini, G.L., 1986, ApJ 300, 77

[13] Huchra, J., Davis, M., Latham, D.,Tonry, J., 1983, ApJS 52, 89

[14] Montuori, M., Labini, F.S., 1997, AJ, 487, L21

[15] Montuori, M., Labini, F.S., Amici, A., 1997, Physica A, 246, 1

[16] Peebles, P.J.E., 1980, The Large Scale Structure of the Universe, Princeton University Press, Princeton, NJ.

[17] Rund, H., 1973, The Hamilton-Jacobi Theory in the Calculus of Variations, Krieger, NY, p. 176

[18] Vettolani, G., et al., in: Proc. of Schloss Rindberg Workshop: Studying the Universe With Clusters of Galaxies

A DERIVATION OF PARTICLE STRUCTURES AND THE DIRAC EQUATION FROM FUNDAMENTAL SYMMETRIES

PETER ROWLANDS

IQ Group, University of Liverpool, Department of Physics, Oliver Lodge Laboratory, Oxford Street, P. O. Box 147, Liverpool, L69 3BX, UK. e-mail prowl@hep.ph.liv.ac.uk

J. P. CULLERNE

IQ Group, University of Liverpool, Department of Computer Science, Chadwick Laboratory, Peach Street, Liverpool, L69 72F, and Winchester College, Winchester, UK. e-mail jpculleme@hotmail.com

ABSTRACT

The broken symmetry responsible for the spectrum of fundamental particles and the characteristic properties of the strong, electromagnetic and weak interactions is derived from a representation of the three source terms or charges, *s*, *e*, *w*, by the quaternions operators, *i*, *j*, *k*. This symmetry is shown to be identical to that required by the Dirac equation in expressing the quantized version of the relativistic relation between energy, mass and momentum, and to be compatible with the simplest SU(5) scheme which has been proposed for Grand Unification. The significant features of the Standard Model are reproduced from basic principles, including the spectrum of masses for baryon and meson states, and a new value for the Weinberg angle.

1 A 32-Part Algebra

The existence of the fundamental particles classified as quarks and leptons and the composite structures that result from them are regularly regarded as the result of a broken symmetry between the electromagnetic, strong and weak interactions. The questions which this representation suggests are: what would be the nature of the unbroken symmetry; and what causes the breaking of it? It may be possible to answer these questions if we begin from even deeper levels of symmetry within fundamental physics. Some of these symmetries stem from the fundamental properties of the algebras used to represent physical quantities.

Many different algebras are used to relate physical quantities, but two particularly significant ones are the 4-vector algebra (with three real parts and one imaginary), which is used as the basis of special relativity, and the mirror-image quaternion system (with three imaginary parts and one real) from which vectors were originally derived. 4-vector algebra is characteristically applied to space and time, the three real vector units **i**, **j**, **k** being applied in the first instance to the three spatial dimensions, and the imaginary scalar unit *i* being applied to time. Quaternion algebra has never found a universal application, partly perhaps for historical reasons, when it became perceived as the losing candidate in the battle for the fundamental representation of three-dimensional physical quantities, and was banished from serious consideration in mainstream physics textbooks. It has, however, found many special applications, and has several appealing features, especially the fact that its dimensionality is fixed, for good mathematical reasons, at exactly three. It has been suggested that a significant application of quaternion algebra may be to the sources of the four fundamental physical interactions, the three imaginary quaternion units *i*, *j*, *k* being applied to the sources of the electromagnetic, strong and weak interactions, collectively described as 'three-dimensional charge', while the real scalar unit 1 is applied to mass, the source of the gravitational interaction.[1,2] Such a representation (existing in its ideal form only at Grand Unification) would suggest why nongravitational forces are repulsive for identical particles while gravitational forces are attractive, and may be a component in a more general symmetry between the parameters space, time, mass and charge. An exact symmetry between a space-time 4-vector and a mass-charge quaternion could also be related to recent developments in multivariate vector algebra, which, following the original rule proposed by Clifford, define the full product of vectors **a** and **b** as

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

in exact analogy with quaternion multiplication.[3,4] Vectors then have the multiplication rules of Pauli matrices, and the full product has the property of closure, unlike the individual scalar and vector products. This multiplication rule will be assumed to apply to all vectors in this discussion.

An interesting result is obtained when we combine the eight units of the 4-vector and quaternion systems (**i**, **j**, **k**, *i*, *i*, *j*, *k*, 1) into a single algebra. This algebra has 32 parts: 1 real scalar, 1 imaginary scalar, 3 real vectors, 3 imaginary vectors, 3 quaternions, 3 imaginary quaternions, 9 real vector quaternions and 9 imaginary vector quaternions.[5,6,7] The existence of 32 parts suggests that it can be generated from a binomial combination of five 'primitive' (though composite) components. An example of such a set of 'primitive' components is ***k***, ***ii***, ***ii*j**, ***ii*k**, ***ij***. All possible units of the algebra emerge as the 32 terms of the binomial product (1 + ***k***) (1 + ***ii*i**) (1 + ***ii*j**) (1 + ***ii*k**) (1 + ***ij***), with + or – signs being determined by the arbitrary ordering of the factors. (A full listing is given in Appendix 1.) Since there is reason to suppose that space, time, mass and charge are the most fundamental set of parameters available to physics,[1,2] we might expect that the 32-part algebra containing their units is of fundamental physical significance, and that significant physics will also be contained in the 5-fold structure represented by the binomial 'primitives'. This is essentially the algebra used in the Dirac equation.

The Dirac algebra may, in fact, also be seen as a result of the competing conditions required for conserved and nonconserved quantities. Essentially, the Dirac equation results from a quantum version of the classical relativistic equation for energy-momentum conservation:

$$E^2 - p^2 = m^2 .$$

Now, normally, we concentrate on the variation of the individual components of the 4-vector (iE, **p**) represented in the left-hand side of the equation. Classically, if we wish to specify the *invariance* of the total quantity $E^2 - p^2$ for an object of fixed mass m, then we need to extend the 4-vector to an expression which specifically incorporates m. Just as the 4-vector representation requires complex algebra, its extension is only possible through a noncommutative (specifically, quaternion) algebra with, say, units such as ***ik**E*, **ip**, ***j**m*.

For a quantum system with stationary states, in which E and **p** are *fixed*, we describe the fixity of E, **p** and m against the variation of the space and time coordinates in the exponential part of the Dirac wavefunction, $e^{-i(Et\,-\,\mathbf{p.r})}$. Remarkably, the structures for particles generated by applying the quaternion operators i, j, k to the charges s, e, w seem to be essentially identical in form to those produced by the Dirac wavefunction in applying i, j, k to the 3-'dimensional' 'mass-operator' (iE, **p**, m). The combination in each case requires a 5-dimensional term which automatically breaks the symmetry that would otherwise apply to the coefficients of the quaternion operators. The 5-dimensionality also suggests a possible link with the SU(5) symmetry which is the simplest possible candidate for Grand Unification.

2 The Dirac Equation, Matter Quaternions and Wave Functions

2.1 The quaternion Dirac operator

A formal development of the Dirac equation into quaternion form is relatively simple to establish.[8,9] The development requires an association of real and imaginary quaternions and vectors with the γ matrices which are used in most standard texts.[10,11,12] With an appropriate association, one converts

$$(\gamma^\mu \partial_\mu + im)\,\psi = 0 \quad (1),$$

where $\mu = 0, 1, 2, 3$, into the following quaternion forms:

$$(i\partial_t + ik\nabla - m)\,\psi = 0 \quad (2a), \text{ or } \quad (ik\partial_t + i\nabla + ij\,m)\,\psi = 0 \quad (2b),$$

where

$$\nabla = \mathbf{i}\partial_x + \mathbf{j}\partial_y + \mathbf{k}\partial_z \quad (3a)$$

is a vector operator and the multiplication of two unit vectors is understood to mean the full product that one might expect from σ matrices; that is,

$$\mathbf{ab} = \mathbf{a.b} + i\,\mathbf{a} \times \mathbf{b} \quad (3b),$$

where the first and second terms are respectively the scalar and vector products.

Here we have used the mapping:

$$\gamma^0 = -ii$$
$$\gamma^1 = \mathbf{i}k$$
$$\gamma^2 = \mathbf{j}k$$
$$\gamma^3 = \mathbf{k}k$$
$$\gamma^5 = ij\,,$$

for (2a), and

$$\gamma^0 = ik$$
$$\gamma^1 = \mathbf{i}i$$
$$\gamma^2 = \mathbf{j}i$$
$$\gamma^3 = \mathbf{k}i$$
$$\gamma^5 = ij\,,$$

for (2b).

The various versions of the Dirac equation may be generated by multiplying by the appropriate quaternion operators. Here for example, (2a) may be obtained from (2b) by multiplying from the left by $-ij$.

The expression (2b) effectively allocates the differential operators which appertain to the parameters E, $\mathbf{p}$, and m to quaternions. This idea becomes invaluable when we consider the effect of C (charge conjugation), P (parity) or T (time reversal) symmetry operations on quaternion expressions. It is for this reason that we choose our quaternion Dirac operator to be the one in (2b), namely

$$(ik\partial_t + i\nabla + ij\, m) \quad (4).$$

2.2 Matter quaternions

Once the Dirac algebra has been established, it is a relatively easy task to show that the Dirac equation follows from a quantization of a basic Lorentz-invariant classical conservation of energy. The 32-part algebra resulting from a combination of 4-vectors and quaternions simplifies to a virtually pure quaternion algebra, with the exception of the operator $i\nabla$ which retains some vector character.

We begin our analysis with the Lorentz-invariant relationship between energy, mass and momentum:

$$E^2 + p^2 + m^2 = 0 \quad (5),$$

where p^2 is the scalar product $\mathbf{p}.\mathbf{p}$. It is a relatively simple matter to factorize this expression using complex and non-commutative algebra, viz. quaternions. There are four different ways of doing this:

$$(kE + ii\,\mathbf{p} + ij\,m)\,(kE + ii\,\mathbf{p} + ij\,m) = 0 \quad (6a),$$
$$(kE - ii\,\mathbf{p} + ij\,m)\,(kE - ii\,\mathbf{p} + ij\,m) = 0 \quad (6b),$$
$$(-kE + ii\,\mathbf{p} + ij\,m)\,(-kE + ii\,\mathbf{p} + ij\,m) = 0 \quad (6c),$$
$$(-kE - ii\,\mathbf{p} + ij\,m)\,(-kE - ii\,\mathbf{p} + ij\,m) = 0 \quad (6d).$$

We could have chosen to incorporate negative mass terms into the quaternion factors. However, to make the factors different we would have to have made either E always positive or $\mathbf{p}$ always positive. Also, work on fundamental parameters and their properties,[1,2,8] suggests that the parameter m is an absolute continuum which cannot support negative terms.

The quaternion factors in (6a) to (6d) are linearly independent; that is, one cannot be written as a linear combination of the others. We label them as follows:

$$q_1 = (kE + ii\,\mathbf{p} + ij\,m) \quad (7a),$$
$$q_2 = (kE - ii\,\mathbf{p} + ij\,m) \quad (7b),$$
$$q_3 = (-kE + ii\,\mathbf{p} + ij\,m) \quad (7c),$$
$$q_4 = (-kE - ii\,\mathbf{p} + ij\,m) \quad (7d).$$

In order to distinguish these quaternions from other quaternion quantities which appear in this paper, we refer to q_1 to q_4 as matter quaternions.

The properties of these matter quaternions under C, P and T operations demonstrate very nicely the way in which the parameters E, $\mathbf{p}$ and m are affected by these fundamental symmetry operations. C, P and T operations have a very simple representation in quaternion formalism. Energy, momentum and mass have been allocated to quaternions in these factors. A reversal of spatial coordinates or parity transformation (P) is clearly associated with the vector $\mathbf{p}$ term, and time reversal (T) with the E term, while a charge conjugation or replacement of particle with anti-particle (C) would involve both E and $\mathbf{p}$, or, alternatively just m. Each of these parameters, E, $\mathbf{p}$, or m is therefore associated with a different fundamental symmetry.

C, P and T operations on, say, the matter quaternion q_1, may then be represented as the following operations:

P: $$-i\, q_1\, i = -i\,(kE + ii\,\mathbf{p} + ij\,m)\, i = (-kE + ii\,\mathbf{p} - ij\,m) = -q_2 \quad (8a),$$
T: $$-k\, q_1\, k = -k\,(kE + ii\,\mathbf{p} + ij\,m)\, k = (kE - ii\,\mathbf{p} - ij\,m) = -q_3 \quad (8b),$$
C: $$-j\, q_1 j = -j\,(kE + ii\,\mathbf{p} + ij\,m) j = (-kE - ii\,\mathbf{p} + ij\,m) = q_4 \quad (8c).$$

It is also apparent that:

CP = T: $$-j\,(-i\, q_1\, i) j = -k\,(kE + ii\,\mathbf{p} + ij\,m)\, k = (kE - ii\,\mathbf{p} - ij\,m) = -q_3 \quad (9a),$$
PT = C: $$-i\,(-k\, q_1\, k)\, i = -j\,(kE + ii\,\mathbf{p} + ij\,m) j = (-kE - ii\,\mathbf{p} + ij\,m) = q_4 \quad (9b),$$
TC = P: $$-k\,(-j\, q_1 j)\, k = -i\,(kE + ii\,\mathbf{p} + ij\,m)\, i = (-kE + ii\,\mathbf{p} - ij\,m) = -q_2 \quad (9c),$$

and that

TCP ≡ identity:
$$-k\,(-j\,(-i\, q_1\, i) j)\, k = -kji\,(kE + ii\,\mathbf{p} + ij\,m)\, ijk = ijk\,(kE + ii\,\mathbf{p} + ij\,m)\, ijk = q_1 \quad (10).$$

The negative signs which accompany the matter quaternions in (8a), (8b), (9a) and (9c) are effectively indetectable phase multipliers. They remind us of the fact that our factorizations in (6a) to (6b) could have been with ± the factors. What is important here is to note what happens to the relative signs of the parameters. Since mass is chosen to be unipolar, we can summarise the results of our quaternion operations above as follows:

Operation P on matter quaternion → reverses sign of $\mathbf{p}$ with respect to E and m, e.g.
$$\mathrm{P}(q_1) \rightarrow q_2 .$$
Operation T on matter quaternion → reverses sign of E with respect to $\mathbf{p}$ and m, e.g.
$$\mathrm{T}(q_1) \rightarrow q_3 .$$
Operation C on matter quaternion → reverses sign of $\mathbf{p}$ and E with respect to m, e.g.
$$\mathrm{C}(q_1) \rightarrow q_4 .$$
Operation TCP on matter quaternion → acts as identity on quaternion, e.g.
$$\mathrm{TCP}(q_1) \rightarrow q_1 .$$

A violation of charge conjugation, such as happens (at least partially) in the weak interaction, would effectively make the parameter m unable to assume a sign opposite to $\mathbf{p}$ and E. Violation of C would therefore lead to something like:

$$-j\,(q_1)\,j \rightarrow q_1 \quad (11).$$

Similar arguments could be applied to define violations of P and T in this formalism.

In such a case, since

$$-j\,(q_1)\,j = -j\,(\boldsymbol{k}E + \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,j = (-\boldsymbol{k}E - \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m) \rightarrow q_1 - 2m\boldsymbol{i}j \quad (12),$$

that part of the matter quaternion which is involved in the weak interaction would require a term like $2m\boldsymbol{i}j$ to be equivalent to zero. The matter quaternion for a neutrino which only has a weak component of charge,[1,9] and involves violation of charge conjugation, would, in the simplest analysis, require a particle of zero mass.

In general, to reduce a term like $\pm\, 2m\boldsymbol{i}j$ to 0 requires the addition of its negative value elsewhere. Weak interactions require such additions for fermions where violation of charge conjugation makes the weak interaction unable to distinguish between the weak charge components $\pm$ w.[1,8]

2.3 Quaternion solutions to the Dirac equation and wave-functions

The factorizations (6a) to (6d) may be used to construct quaternion solutions of the Dirac equation in (2b). The Dirac operator in quaternion form (4) may be considered as acting on quaternion wavefunctions, generating four linearly independent ways of writing (4) which coincide with the four factorizations (6a) to (6d). The four quaternion functions:

$$(\boldsymbol{k}E + \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,e^{-i(Et-\mathbf{p}.\mathbf{r})} \quad (13a),$$
$$(\boldsymbol{k}E - \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,e^{-i(Et+\mathbf{p}.\mathbf{r})} \quad (13b),$$
$$(-\boldsymbol{k}E + \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,e^{i(Et+\mathbf{p}.\mathbf{r})} \quad (13c),$$
$$(-\boldsymbol{k}E - \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,e^{i(Et-\mathbf{p}.\mathbf{r})} \quad (13d).$$

with $\mathbf{p} = \mathbf{i}p_x + \mathbf{j}\,p_y + \mathbf{k}\,p_z$, are all solutions of (2b), each satisfying the equation by forming one of the factorizations (6a) to (6d); e.g.

$$(i\boldsymbol{k}\partial_t + \boldsymbol{i}\nabla + \boldsymbol{i}j\,m)\,(\boldsymbol{k}E + \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,e^{-i(Et-\mathbf{p}.\mathbf{r})} = (\boldsymbol{k}E + \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m)\,(\boldsymbol{k}E + \boldsymbol{i}i\,\mathbf{p} + \boldsymbol{i}j\,m) = 0.$$

By making a mapping between the γ matrices and multivariate vector quaternions, we were able to avoid any matrix manipulation in the solution of the Dirac equation. However, the fact that there can be no privileging of sign for quaternions leads us once again to four linearly independent solutions; a result that one would have obtained if one had originally considered the γ matrix form.

In order to see this idea more clearly and to obtain a set of orthogonal eigenstates, let us at this juncture take our considerations a step further back. The set of functions:

$$e^{-i(Et-\mathbf{p}.\mathbf{r})},\ e^{-i(Et+\mathbf{p}.\mathbf{r})},\ e^{i(Et+\mathbf{p}.\mathbf{r})},\ e^{i(Et-\mathbf{p}.\mathbf{r})},$$

are eigenfunctions of the operator $(\boldsymbol{ik}\partial_t + \boldsymbol{i}\nabla + \boldsymbol{ij}\ m)$ with the corresponding eigenvalues:

$$(\boldsymbol{k}E + \boldsymbol{ii}\ \mathbf{p} + \boldsymbol{ij}\ m),\ (\boldsymbol{k}E - \boldsymbol{ii}\ \mathbf{p} + \boldsymbol{ij}\ m),\ (-\boldsymbol{k}E - \boldsymbol{ii}\ \mathbf{p} + \boldsymbol{ij}\ m), \text{ and } (-\boldsymbol{k}E - \boldsymbol{ii}\ \mathbf{p} + \boldsymbol{ij}\ m).$$

Using the bra-ket notation we could write the eigenvalue equations for the operator:

$$\mathcal{D} = (\boldsymbol{ik}\partial_t + \boldsymbol{i}\nabla + \boldsymbol{ij}\ m)$$

as

$$\mathcal{D}\,|\psi_1\rangle = q_1\,|\psi_1\rangle\,,$$
$$\mathcal{D}\,|\psi_2\rangle = q_1\,|\psi_2\rangle\,,$$
$$\mathcal{D}\,|\psi_3\rangle = q_1\,|\psi_3\rangle\,,$$
$$\mathcal{D}\,|\psi_4\rangle = q_1\,|\psi_4\rangle\,,$$

where $|\psi_1\rangle = e^{-i(Et-\mathbf{p}.\mathbf{r})}\,|1\rangle$, $|\psi_2\rangle = e^{-i(Et+\mathbf{p}.\mathbf{r})}\,|2\rangle$, $|\psi_3\rangle = e^{i(Et+\mathbf{p}.\mathbf{r})}\,|3\rangle$, $|\psi_4\rangle = e^{i(Et-\mathbf{p}.\mathbf{r})}\,|4\rangle$, and $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle,|\psi_4\rangle\}$ are components of a basis (not necessarily orthonormal) to be determined.

One may find the Hermitian conjugate for $\mathcal{D}$ by appealing to the mappings for γ matrices which lead to the construction of $\mathcal{D}$ in quaternion form. We find that the Hermitian conjugate of $\mathcal{D}$ is easily obtained by making a complex-quaternion conjugation; that is, conjugating imaginary i as well as $\boldsymbol{i},\boldsymbol{j}$ and $\boldsymbol{k}$. The operator $\mathcal{D}^\dagger$ is the hermitian conjugate of $\mathcal{D}$ where $\mathcal{D}^\dagger$ is:

$$\mathcal{D}^\dagger = (\boldsymbol{ik}\partial_t - \boldsymbol{i}\nabla + \boldsymbol{ij}\ m)\,.$$

Using this we may now construct matrix elements for the operator $\mathcal{D}$ in the basis $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle,|\psi_4\rangle\}$. The matrix element $\langle\psi_2|\ \mathcal{D}\ |\ \psi_1\rangle = q_1\langle\psi_2|\psi_1\rangle$, may also be constructed by applying $\mathcal{D}^\dagger$ to $|\psi_2\rangle$ first; that is, $\langle\psi_2|\ \mathcal{D}\ |\ \psi_1\rangle = (\mathcal{D}^\dagger|\psi_2\rangle)^\dagger|\psi_1\rangle = q_3\langle\psi_2|\psi_1\rangle$. This means that $q_1\langle\psi_2|\psi_1\rangle = q_3\langle\psi_2|\psi_1\rangle$, which can only be true if $\langle\psi_2|\psi_1\rangle = 0$; that is, $|\psi_1\rangle$ and $|\psi_2\rangle$ are orthogonal. Similar calculations for $\langle\psi_3|\ \mathcal{D}\ |\ \psi_1\rangle$ and $\langle\psi_4|\ \mathcal{D}\ |\ \psi_1\rangle$ lead one to the conclusion that $\{|\psi_1\rangle,|\psi_4\rangle\}$ form a subspace that is necessarily orthogonal to $\{|\psi_2\rangle,|\psi_3\rangle\}$. $|\psi_1\rangle$ and $|\psi_4\rangle$ are not necessarily orthogonal and the same goes for $|\psi_2\rangle$ and $|\psi_3\rangle$. The most general basis one could write for $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle,\ \psi_4\rangle\}$ is:

$$|\psi_1> = \begin{pmatrix}\alpha_1\\0\\0\\\beta_1\end{pmatrix} e^{-i(Et-\mathbf{p}.\mathbf{r})},\ |\psi_2> = \begin{pmatrix}0\\\alpha_2\\\beta_2\\0\end{pmatrix} e^{-i(Et+\mathbf{p}.\mathbf{r})},\ |\psi_3> = \begin{pmatrix}0\\\alpha_3\\\beta_3\\0\end{pmatrix} e^{i(Et+\mathbf{p}.\mathbf{r})},\ |\psi_4> = \begin{pmatrix}\alpha_4\\0\\0\\\beta_4\end{pmatrix} e^{i(Et-\mathbf{p}.\mathbf{r})}.$$

However, with further consideration we find that these eigenfunctions may be orthogonalized using a Gram-Schmidt method leading to a basis which we will use throughout this work of the following form:

$$|\psi_1> = \begin{pmatrix}1\\0\\0\\0\end{pmatrix} e^{-i(Et-\mathbf{p}.\mathbf{r})}, |\psi_2> = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} e^{-i(Et+\mathbf{p}.\mathbf{r})}, |\psi_3> = \begin{pmatrix}0\\0\\1\\0\end{pmatrix} e^{i(Et+\mathbf{p}.\mathbf{r})}, |\psi_4> = \begin{pmatrix}0\\0\\0\\1\end{pmatrix} e^{i(Et-\mathbf{p}.\mathbf{r})}.$$

Since the application of $\partial\!\!\!/$ onto the above states leads to solutions of the Dirac equation $(\boldsymbol{ik}\partial_t + \boldsymbol{i}\nabla + \boldsymbol{ij}\, m)\, \psi = 0$, we find that the quaternion wave functions ψ_1, ψ_2, ψ_3, ψ_4, are given by:

$$(\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, |\psi_1\rangle \quad (14a),$$
$$(\boldsymbol{k}E - \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, |\psi_2\rangle \quad (14b),$$
$$(-\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, |\psi_3\rangle \quad (14c),$$
$$(-\boldsymbol{k}E - \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, |\psi_4\rangle \quad (14d).$$

If we consider an alignment of the momentum vector $\mathbf{p}$ in the direction of the z axis (p_x and p_y are both zero), we see that $\pm\,\mathbf{p}$, which appears in our states (14a) to (14d), may be used to identify the eigenstates of helicity. Our convention has been chosen to coincide with the normal Dirac conventions for positive and negative energy states and spin orientations. (14a) represents an electron with spin parallel to momentum, (14b) an electron with spin anti-parallel to momentum, (14c) represents a positron with spin parallel to momentum, (14d) a positron with spin anti-parallel to momentum.

This is consistent with what happens to these states on application of T, C or P. Indeed, now we have a basis within which we can construct operators in matrix form, the matrix operators for T, P or C are as follows:

$$T = \begin{pmatrix} 0 & 0 & e^{2iEt} & 0 \\ 0 & 0 & 0 & e^{2iEt} \\ e^{-2iEt} & 0 & 0 & 0 \\ 0 & e^{-2iEt} & 0 & 0 \end{pmatrix}, P = \begin{pmatrix} 0 & e^{-2ipz} & 0 & 0 \\ e^{-2ipz} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{2ipz} \\ 0 & 0 & e^{2ipz} & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} 0 & 0 & 0 & e^{2i(Et-pz)} \\ 0 & 0 & e^{2i(Et+pz)} & 0 \\ 0 & e^{-2i(Et+pz)} & 0 & 0 \\ e^{-2i(Et-pz)} & 0 & 0 & 0 \end{pmatrix}$$

2.4 Annihilation and creation operators

We have seen that the quaternion operator $(\boldsymbol{ik}\partial_t + \boldsymbol{i}\partial_x + \boldsymbol{ij}\, m)$ acting on a state

$$(\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, e^{-i(Et - \mathbf{p.r})}$$

leads to the equation: $(\boldsymbol{ik}\partial_t + \boldsymbol{i}\partial_x + \boldsymbol{ij}\, m)\, (\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, e^{-i(Et - \mathbf{p.r})} = 0$. This operation with operator $\not{\partial}$ may be thought of as a creation operation acting on the single particle fermion state which is already filled. The result is therefore zero. We may obtain the corresponding annihilation operation by finding the Hermitian conjugate of $(\boldsymbol{ik}\partial_t + \boldsymbol{i}\partial_x + \boldsymbol{ij}\, m)$.

We know:

$$\not{\partial}\, e^{-i(Et - \mathbf{p.r})} = q_1\, e^{-i(Et - \mathbf{p.r})} \qquad (15)$$

where $\not{\partial}$ is acting as though it were the creation operator $\mathbf{a}^\dagger$ acting on the group of translations and rotations that we call vacuum. Incorporating a normalization factor, we can therefore write $\mathbf{a}^\dagger$ as

$$\mathbf{a}^\dagger = (1/2E)\, (\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\,.$$

Now, the Hermitian conjugate of expression (15) is:

$$\not{\partial}^\dagger\, e^{-i(Et - \mathbf{p.r})} = q_3\, e^{i(Et - \mathbf{p.r})} \qquad (16)$$

Here, because it describes the creation of the anti-particle, $\not{\partial}^\dagger$ is acting as though it were the annihilation operator a conjugate to $\mathbf{a}^\dagger$, so that

$$\mathbf{a} = (1/2E)\, (-\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\,.$$

a, like $\mathbf{a}^\dagger$, has been modified by a factor $(1/2E)$ so that, when acting on a state, $\mathbf{a}^\dagger\mathbf{a}$ will reproduce the state up to a scalar multiplicative factor which does not affect the quaternion properties of the state.

It is easy to verify that these two operators have the commutation relations appropriate to fermion annihilation and creation operators:

$$\mathbf{a}\mathbf{a}^\dagger + \mathbf{a}^\dagger\mathbf{a} = 1\,.$$

We now need to find the vacuum state that $\mathbf{a}^\dagger$ can act upon to lead to the single particle state:

$$(\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, e^{-i(Et - \mathbf{p.r})}\,.$$

We need only consider the following:

$$\mathbf{a}\, (\boldsymbol{k}E + \boldsymbol{ii}\,\mathbf{p} + \boldsymbol{ij}\, m)\, e^{-i(Et - \mathbf{p.r})} = \text{VAC}$$
$$= 2E\, (1/2E)\, (E - \boldsymbol{ij}\,\mathbf{p} + \boldsymbol{ii}\, m)\, e^{-i(Et - \mathbf{p.r})}\,.$$

To check this we need only consider

$$\mathbf{a}^{\dagger}\,(\mathbf{a}\,(kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et-\mathbf{p.r})} = \mathbf{a}^{\dagger}\,\text{VAC}$$
$$= \mathbf{a}^{\dagger}\,(1/2E)\,(2E\,(E - ij\,\mathbf{p} + ii\,m)\,e^{-i(Et-\mathbf{p.r})}$$
$$= (1/2E)\,2E\,(kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et-\mathbf{p.r})}$$
$$= (kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et-\mathbf{p.r})}\,.$$

It is easy to verify that the further action of **a** onto the VAC state leads to zero.

2.5 Bilinear covariants

In finding the bilinear covariants, it will be convenient to use the form of the wave function suggested by the mapping of gamma matrices represented in (2a), as this is closest to that conventionally used. Here,

$$\psi = (ii\,E + k\,\mathbf{p} - m)\,e^{-i(Et-\mathbf{p.r})}\,,$$

and the more usual form can be recovered by multiplying from the right by $-ij$. The Hermitian conjugate now becomes the complex quaternion conjugate of this expression:

$$\psi^{\dagger} = (ii\,E - k\,\mathbf{p} - m)\,e^{i(Et-\mathbf{p.r})}\,,$$

leading to the Hermitian conjugate equation:

$$-i\partial_t\,\psi^{\dagger}\,\gamma^{o}\; - i\nabla\,\psi^{\dagger}(-\gamma) - m\,\psi^{\dagger} = (ii\,E - k\,\mathbf{p} - m)\,e^{i(Et-\mathbf{p.r})}\,(-ii\,E + k\,\mathbf{p} - m) = 0\,.$$

The adjoint wave function take the form

$$\bar{\psi} = \psi^{\dagger}\,\gamma^{o} = \psi^{\dagger}\,(-ii) = (-E + ij\,\mathbf{p} + ii\,m)\,e^{i(Et-\mathbf{p.r})}\,,$$

leading to the adjoint equation:

$$i\partial_t\,\bar{\psi}\,\gamma^{o}\; + i\nabla\;\bar{\psi}\,\gamma + m\;\bar{\psi} = (-E + ij\,\mathbf{p} + ii\,m)\,e^{i(Et-\mathbf{p.r})}\,(ii\,E + k\,\mathbf{p} + m) = 0\,.$$

From these expressions we may derive the probability and current densities and the other bilinear covariants, summing over the terms representing the four solutions ψ_1, ψ_2, ψ_3, ψ_4, as in the conventional theory, and dividing by the normalization term $8E^2$:

$$\bar{\psi}\,\psi = -\,ii\,m^2\,/\,E^2$$
$$\underline{\psi}\,\gamma^{o}\psi = \psi^{\dagger}\psi = 1$$
$$\underline{\psi}\,\gamma\psi = -\,ij\,p^2\mathbf{1}\,/\,E^2$$
$$\underline{\psi}\,\gamma^{5}\psi = 0$$
$$\underline{\psi}\,\gamma^{5}\gamma\psi = -\mathbf{1}$$
$$\underline{\psi}\,\gamma^{5}\gamma^{o}\psi = ij\,p^2\mathbf{1}\,/\,E^2$$
$$\underline{\psi}\,\sigma^{mn}\psi = im^2\mathbf{1}\,/\,E^2$$
$$\psi\,\sigma^{on}\psi = 0\;\;.$$

3 Some Fundamental Symmetries

3.1 Conservation and nonconservation

The quaternion representation of mass and charge may be considered symmetrical to the more familiar 4-vector representation of space and time, but this is not the only symmetry relating these quantities. Another is the relationship between conserved and nonconserved quantities.[1,13,14] Mass and charge are conserved quantities while space and time are nonconserved. Now, physics is based on defining systems in which conserved quantities remain fixed while nonconserved quantities vary absolutely. A conserved quantity can only be defined with respect to changes in a nonconserved quantity. In effect, we look at how mass and charge, and such quantities as energy, momentum, force or action remain constant, or zero, or a maximum or a minimum, because of the more fundamental requirements involving mass and charge, while the space and time coordinates alter arbitrarily.

Also, the conservation laws of mass and charge are not merely global, applying to the total amount of each quantity in the universe, but, more exactly, local, applying to the amount of each quantity at a given place in a given time. Exactly the same applies to nonconservation, which is, in every respect, the exact opposite of conservation and just as definite a property. Just as the elements of mass and charge have individual, specific and permanent identities, so those of space and time have no identity whatsoever, a property which has a number of different physical manifestations. Space and time, for example, both exhibit translation symmetry. Every element of space and time is exactly like every other, and is not only indistinguishable in practice, but must be stated to be indistinguishable when we write down physical equations.

The connections between conservation laws and fundamental symmetries have been determined by a number of mathematical approaches, using powerful and sophisticated results such as Noether's theorem and Weyl's principle of gauge invariance. Yet the existence of these fundamental symmetries does not depend on the knowledge of such mathematical approaches, but is present at a more basic level in which conserved and nonconserved quantities take symmetrically opposite properties. According to Noether's theorem, conservation laws in physics are associated with fundamental symmetries – every continuous symmetry of a Lagrangian requires a quantity which is conserved by its dynamics; but this is essentially because conserved quantities need nonconserved ones to define them.

Noether's theorem shows us that the translation symmetry of time is precisely identical to the conservation of energy, and that the translation symmetry of space is precisely identical to the conservation of linear momentum. This is significant, but, more significant for our purposes is the fact that space, in addition, because it is three-dimensional, also has rotation symmetry; this means that there is no identity, either, for spatial directions, and Noether's theorem relates this to the conservation of angular momentum. The 'wave' term in the Dirac wavefunction, $e^{-i(Et - \mathbf{p.r})}$, is simply a representation of the group of such space and time translations and rotations. Translation and rotation symmetries, however, are merely representations of the more fundamental property of nonconservation, and nonconservation in one parameter presupposes an exactly opposite form of conservation in another. Thus the Dirac equation uses the translation-rotation properties of the 'wave' term in conjunction with differential operators explicitly describing the nonconservation of space and time to establish the opposing concept of energy-momentum-mass conservation.

We can illustrate the exactly opposite nature of conservation and nonconservation by expressing the identity or uniqueness properties of the conserved quantities mass and charge in terms of 'translation' and 'rotation' *a*symmetries. Translation asymmetry then means that one element of mass or charge cannot be 'translated to' (or exchanged for) any other within a system, however similar; but charge, because it is 'three-dimensional, as well as conserved, must have the further property of rotation asymmetry, meaning that the three types of charge cannot be rotated or converted into each other.

The translation asymmetry of charge becomes significant when we attempt to map the three charge 'dimensions' s, e, w onto the cyclic or rotation symmetric quaternion units $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, because the noncyclic conservation law of charge must then be reconciled with the cyclic nature of the algebraic operators used to represent its units. This, we believe, is the origin of fundamental particle structures, while the Dirac equation similarly results from the mapping of a cyclic quaternion algebra onto a semi-fixed system of energy, momentum and mass. Both the Dirac equation and the structures of particles, then, can be said to originate in the combination of the conditions needed to accommodate systems which are cyclic with those which are not.

The absoluteness of the nonconservation properties of space and time becomes apparent in the gauge invariance used in both classical and quantum electrodynamics. Electric and magnetic field terms remain invariant under arbitrary changes in the vector and scalar potentials, or phase changes in the quantum mechanical wavefunction, resulting from what are essentially translations (or rotations) in the space and time coordinates. A gauge invariant system is one that remains conservative under arbitrary changes in the coordinates which do not result in changes in the values of such conserved quantities as charge, energy, momentum and angular momentum. In principle, we cannot know the absolute phase or value of potential because we cannot choose to fix values of coordinates which are subject to absolute and arbitrary change. Of special significance is the fact that, in the Yang-Mills principle used in particle physics, the arbitrary phase changes are specifically local, rather than global. Nonconservation, therefore, must be local in exactly the same way as conservation.

3.1 Continuity and discontinuity

Another important symmetry relates continuous or noncountable quantities, which include mass and time, with discontinuous, discrete or countable ones, which include space and charge. Both concepts (the 'analogue' and 'digital') exist in nature because they are fundamental properties of the fundamental parameters. The divisibility of space is essential to physics, because space is always used in direct measurement. It is impossible, in fact, to measure anything else – 'time'-measuring devices, for instance, all use some method of repetition of a spatial interval. The whole process of measurement depends crucially on the divisibility of space, or creation of discontinuities within it. Space, also, is reversible, unlike time, and it is this reversibility of space (the repetition of spatial intervals) which is used in the measurement of time.

Space can be discontinuous in both quantity and direction; in addition to being reversed, iit can also be changed in orientation; and, without both of these properties, measurement would be impossible. Time is irreversible because it cannot actually be divided. It cannot be reversed precisely because it is absolutely continuous, and for the same reason it cannot be multidimensional in the same way as space. Any reversal of time or change of orientation would require discontinuity. The same distinction occurs between mass and charge. Mass, in its most general form as a gravitational source, is an absolute continuum present in all systems and at every point in space; this is why there is no negative mass, for negative mass would necessarily require a break in the continuum, and it is also why mass exists in only one kind (or 'dimension'). Charge, on the other hand, is divisible and measured in units, though, because charge is a conserved quantity, unlike space, these units must be fixed. Again, charge as a noncontinuous quantity is also dimensional, like space, but unlike the continuous quantity mass.

A major problem in contemporary physics is 'reversibility paradox', where time, according to the laws of physics, is reversible in mathematical sign, when it is clearly not reversible in physical consequences. Time, however, is characterised by imaginary numbers, and imaginary numbers are not privileged according to sign. Thus, time may be expected to have equal positive and negative mathematical solutions as an imaginary mathematical parameter, but it can have only one physical direction as a continuous physical quantity. The corresponding unipolarity, or single sign, of mass is the reason why we have a CPT, rather than an MCPT, theorem, C standing for charge conjugation, P for space reflection and T for time reversal, all of which have two mathematical sign options. The M operator could be considered equivalent to an identity transformation on q_1 like the combined TCP.

The distinction between continuous and discontinuous quantities has real physical consequences when we mathematically combine space and time in Minkowski's 4-vector representation, as symmetry apparently requires us to do. The mathematical requirement gives us two physical options: we can either make time space-like (or discrete) or space time-like (or continuous). This seems to be the origin of wave-particle duality, the discrete options leading to particles, special relativity and Heisenberg's quantum mechanics, while the continuous options lead to waves, Lorentzian relativity and Schrödinger's wave mechanics. The Heisenberg formulation makes everything discrete, including quantities which are normally continuous, so mass becomes charge-like quanta in quantum mechanics; the Schrödinger formulation, on the other hand (which carries over into that of Dirac), makes everything continuous, so charge becomes mass-like wavefunctions in wave mechanics. In measurement, the true situations are restored: Heisenberg reintroduces continuous mass via the uncertainty principle and the virtual vacuum, while Schrödinger reintroduces discrete charge via the collapse of the wavefunction. These distinctions will be necessary to relating charge structures to wave functions.

4 Charge Accommodation

The proposition that the space-time 4-vector may be mapped to a quaternion composed of mass and charge, leads to an interesting problem of charge accommodation. For if mass and charge are to be properly represented, their 'rotational asymmetry' must be accounted for in any quaternion description.

Grand Unified Theories suggest that there is some underlying symmetry between the electromagnetic, strong and weak forces of nature; that is, the effects of these three forces may be made identical when viewed under some particular set of ideal conditions. It is therefore natural to assume that the ultimate sources of the strong and weak interactions could also be terms of the same form as electromagnetic charge.

It has been previously proposed[1] that the unit source for fundamental interactions in physics may be represented as a quaternion, where the unit sources of electromagnetism, strong and weak take on the imaginary part, and the real part is associated with gravity (mass). It is interesting to note that this proposition actually suggests a reason for the peculiar fact that there are four, and only four, fundamental interactions. The employment of imaginary sources ensures that, if there are not exactly two fundamental interactions, with sources described by ordinary complex algebra, then there must be exactly four, with sources described by quaternions.

Two other properties of charge are important and derive from the deeper symmetry[1] linking mass and charge with space and time – here, we will take them to be the fundamental axioms:

- that charges of all kinds exist only as fixed units, and
- that each type of charge is separately conserved in all interactions.

Here we make a distinction between mass and charge even though mass is effectively the charge source of the gravitational interaction. In this section we are interested in the charge accommodation rules which govern the allocation of strong, electromagnetic and weak charge to complex quaternion components.

Let us suppose that charge exists only as fixed units, and that the three types of charge are referred to a 'components':

(s) strong
(e) electromagnetic
(w) weak

These charge components are not interconvertible; that is, they are subject to separate conservation laws. Charge units are fixed, so the only possible values for an individual charge component at any point are 1 and 0; however, the charge units are imaginary, and so + and – values have equal status. Thus a unit charge may be defined as some combination

$$\pm i\text{s},\ \pm j\text{e},\ \pm k\text{w},$$

where

$$\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k},$$

are imaginary quaternion components and *s*, *e* and *w* may take values 1 or 0. The only problem with this is that the charge components *s*, *e* and *w* are fixed with respect to each other and so may be said to have 'rotational asymmetry', whereas the quaternion components ***i***, ***j*** and ***k*** are not fixed and have rotational symmetry. Thus the charge component *e* could be associated with ***i*** or ***k*** as easily as it could be associated with ***j***.

Physical requirements make this seem impossible at first. How can we model a system of entities which do not possess rotational symmetry on a system of entities which do?

The answer to this is quite simple and forces us to consider combinations of unit charges. For example, if we always associate unit values of *e* with the term ***j****e* and zero values of *e* with the terms ***i****e* and ***k****e*, then as long as physical systems with 1***j****e* are indistinguishable from those with 0***i****e* and 0***k****e*, a valid model may be constructed. This would be impossible if unit charges exist independently but it would be possible if unit charges could only exist in some form of combination. Individual charges could then be identified but only in such a way as to never be separable.

Let us try this. Consider a system in which *e* could be identified with any of the quaternion components ***i***, ***j*** and ***k*** but would always be unit when identified with ***j*** and zero when identified with ***i*** or ***k***. Then our charge accommodation system must make combinations of unit charges with 1***j****e* indistinguishable from those containing 0***i****e* and 0***k****e*. Two possible methods are immediately obvious:

- Unit charges could be combined in groups of three in such a way as to render systems containing 1***j****e*, 0***i****e* or 0***k****e* indistinguishable.
- Unit charges could always be combined with unit anti-charges.

It has been suggested[1] that this could be the origin of the coloured quark system and of the combination of quarks in threes or pairs of quark and antiquark. It could therefore be said that something like the experimentally discovered coloured quark system would seem to be required in order to fit quaternions to the parameter charge.

The most obvious way to apply a varying pattern of quaternion designations to *s*, *e* and *w* charge components, would be in the form of a scheme with perfect symmetry between the three charge types, typically

	B	G	R
$\pm e$	***j***	***i***	***k***
$\pm s$	***i***	***k***	***j***
$\pm w$	***k***	***j***	***i***

where the colour labels B(lue), G(reen) and R(ed) have been chosen as the alternative states. Values 1 and 0 may now be allocated as follows:

	B	G	R
$\pm e$	1***j***	0***i***	0***k***
$\pm s$	1***i***	0***k***	0***j***
$\pm w$	1***k***	0***j***	0***i***

This particular designation collapses into a single defined state:

$$\pm s\,i, \pm e j, \pm w j\,.$$

It is exactly this kind of definition we wish to avoid. In such a state s, e and w are always specified to i, j and k respectively; something that cannot be allowed because we must be free to allocate s, e and w to any of the complex quatemions without specifying a particular allocation.

Other patterns which are not perfectly symmetric in s, e and w must therefore be considered. Patterns like

	B	G	R
$\pm e$	$1j$	$1j$	$0i$
$\pm s$	$1i$	$0k$	$0j$
$\pm w$	$1k$	$0i$	$0k$

allow one to allocate charge components to quaternions without ever collapsing the specification to one defined state. The reader will immediately notice that any such scheme automatically introduces asymmetry into the status of the three charge components. In the example above, the s component may take unit value in the B colour only, whereas the e component may take unit value in either B or G, or both. We find that in every conceivable arrangement of any workable scheme involving three 'colour' states, there is always at least one charge component that is confined to a unit value in only one of the colours.

There are eight possible combinations of $\pm e$, $\pm s$, $\pm w$:

$$\begin{array}{ll} +e+s+w & -e-s-w \\ -e+s+w & +e-s-w \\ +e+s-w & -e-s+w \\ -e+s-w & +e-s+w \end{array}$$

Four of these will be antistates of the others. Apart from this distinction, the physical significance of which has still to be determined, the mathematical interchangeability of positive and negative imaginary numbers means that anything which applies to one combination must apply to all the others.

The combinations which correspond to antistates may be determined as soon as we decide which of the three components may take a unit value in only one colour. If every combination of three quarks must contain one of each colour, then the unit component in one colour must always be unit in that colour and must always bear the same sign. If this were not the case it would be possible to identify the particular colours belonging to particular quarks in a particular combination, in violation of the basic principle of the coloured quark system.

The choice is arbitrary because *e*, *s* and *w* up to this stage are only labels, but, by comparison with experiment, we choose *s*. Thus

$$+e+s+w$$
$$-e+s+w$$
$$+e+s-w$$
$$-e+s-w$$

are quarks, to which we may assign the 'flavours' u, d, c, s and

$$-e-s-w$$
$$+e-s-w$$
$$-e-s+w$$
$$+e-s+w$$

are antiquarks with corresponding 'antiflavours' $\bar{u}$, $\bar{d}$, $\bar{c}$ and $\bar{s}$.

With these flavour designations we have effectively accommodated the *s* charge component.

However, such an arrangement is only possible for one of the three charge components because the four identifiable flavours with +*s* require states which have both positive and negative components in *e* and *w*. There is one way of accommodating charge components of opposite sign into a workable scheme of quaternion assignments. We take, for example, the B and G colours of all quarks with the positive charge component to have unit value, and the R component to have zero; at the same time, the R colour of all quarks with negative charge component has unit value, while the B and G colours have zero. Once again, the choice is arbitrary between *e* and *w*, but experimentally, it is the electromagnetic charge component which is incorporated in this way, so that typically the u and d flavours have the following assignments for the *e* and *s* components:

		B	G	R
u	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$0k$	$0j$
	$+w$	k	i	k
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0j$	$1i$
	$+w$	k	i	k

The *w* components have yet to be accommodated. However, it turns out that this is impossible in any way in which the individual colours of the quarks in any combination of three remain unidentified as required, *unless we assume that for the w component of charge, states of + and – are, in some sense, indistinguishable.*

Accommodating weak charge therefore leaves us with a problem. The weak interaction is required to provide a mechanism for making $+w$ and $-w$ indistinguishable. The result of this is the breaking of charge-conjugation symmetry, which requires positive and negative charge components to be indentifiable by the signs of the parity and time-reversal operators. If full CPT symmetry is to be maintained, then either parity or time-reversal symmetry (not both) must also be broken. There are therefore three ways of accommodating w:

- for $+w$ the component may be introduced just like s,
- for $-w$, in order that + and − are indistinguishable, either it is introduced into the table with the proviso that P symmetry is broken, or it is introduced into the table with the proviso that T symmetry is broken.

It has been suggested[1] that this need for three accommodation rules of weak charge leads to the three generations of quark flavour. u and d flavours accommodate $+w$ components, c and s flavours $-w$ components, t and b duplicate the structures of c and s since they represent the two degrees of freedom (P, T) available for producing the required charge conjugation in $-w$; i.e. there is, in effect, no such thing as $-w$, in a particle state, but there are two ways of eliminating it.

Incorporating all these facts into a single set of tables leads to five main representations which we label here A, B, C, D and E. These tables are given in Appendix 2. (A simplified matrix representation, produced by removing the quaternion operators, elucidates only the important information appertaining to the positions of the non-zero quaternion charge components in the allocations.)

In these tables, the simultaneous violation of charge conjugation and parity in the second or (c, s) generation, is symbolized by the term $-z_P w$ for the weak charge component; while the simultaneous violation of charge conjugation and time reversal symmetry in the third or (t, b) generation is symbolized by the term $-z_T w$. Both $-z_P w$ and $-z_T w$ are algebraically equivalent to $+w$, though deriving originally from a requirement for $-w$.

The tables and matrices only implicitly represent the three ways of accommodating w with the three generations of quark flavour, but an algebraic representation can be derived from which the unit and zero charges in tables A-E arise completely naturally, and the reason for exactly five representations then becomes apparent. Let us use the symbols $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ to represent three unit vector components, **i**, **j** or **k**, selected randomly and independently, and let these now be applied to the quaternion elements − j, i, k, representing the units of the three charge components e, s, w (the negative sign for j being selected purely to accord with the historical choice of signs for electrically charged objects). The random unit vector components effectively give us the three degrees of freedom our charge allocation rules require. The expression $(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3)$ then gives us the charge allocation for the d quark. Taking the scalar products of this expression with the respective unit vector components, **i**, **j** and **k**, introduces the idea of quark colour and the three expressions

$$(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{i}$$
$$(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{j}$$
$$(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{k}$$

become the charge allocations for the R, B and G quarks in a ddd baryon, the total (before normalization) being given by $(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{1}$. Whatever values of $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are used, there are only five independent solutions for these scalar products, and these are the ones represented in tables A-E. (The ordering of **i**, **j**, **k** is, of course, completely arbitrary and merely represents the selection of names for the colour labels.)

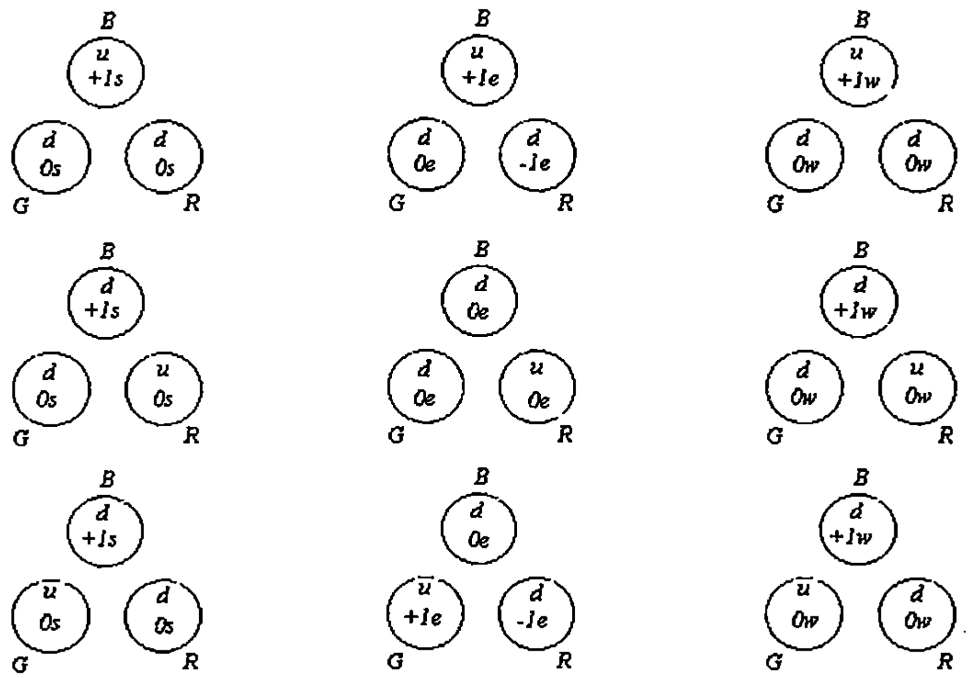
Restricting the quaternion elements to a single sign maintains the perfect symmetry between them, but the symmetry is broken when the degrees of freedom represented by opposing signs are introduced. For the strong charge (represented by i), we introduce antiparticles, by reversing all signs of charge. For the electric charge (represented by j), we alter the sign by adding a unit vector **1** to $\mathbf{r}_1$, thus preserving the randomness of $\mathbf{r}_1$ as required. This produces the transitions within a generation (u/d, c/s, t/b) known as weak isospin. For the weak charge (represented by k), we introduce charge conjugation violation and the factors z_P and z_T, which are responsible for the transitions between what now become the three isospin generations. In each case, the randomness of the vector component $\mathbf{r}_1$, $\mathbf{r}_2$ or $\mathbf{r}_3$, and unrestricted variation over three degrees of freedom, is maintained, and the zero charges arise as naturally from the random scalar products as the units. The charge allocation rules can then be represented in terms of the annihilation and creation operators previously defined, with the full variation specified by the unit vectors.

5 Baryons, Mesons and Leptons

We may obscure quaternion-charge assignments by always combining unit charges in groups of three in such a way as to never know whether we have, for example, $1je$, $0ie$ or $0ke$. This we suspect is the origin of the baryon.

If we remain within the A representation, the diagrams below correspond to the different ways of accommodating charge in a udd combination; that is, the neutron. Each row of these diagrams corresponds to a different quark-colour designation which are of course indistinguishable, in accordance with the basic principle of the coloured quark system.

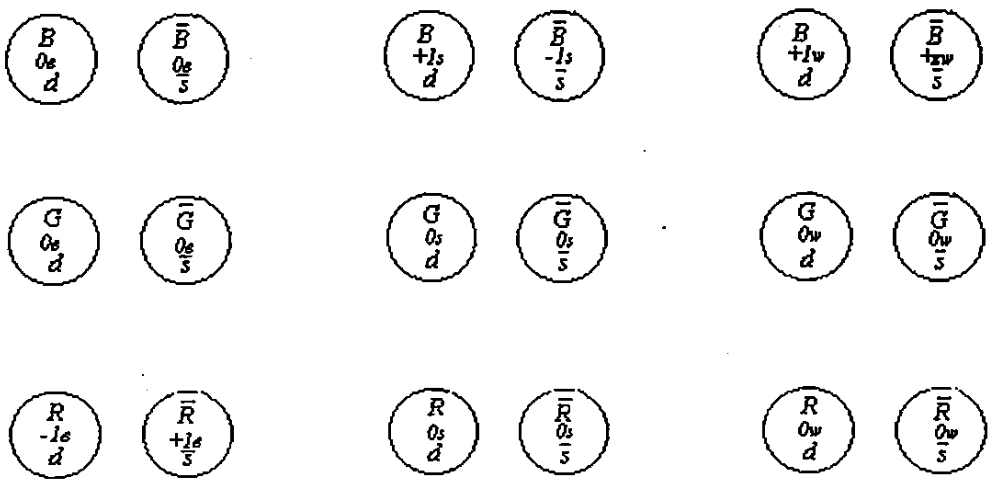
As we can see, the different colour permutations give the same charge structure and at the same time hide the quaternion assignments, e.g. unit je are indistinguishable from zero ie or ke.



Accommodating strong *Accommodating electric* *Accommodating weak*

The obscuring of quaternion assignments can also be accomplished by combining unit charges with unit anticharges. This we suspect is the origin of the meson.

If we remain within the A representation, the diagrams below correspond to the different ways of accommodating charge in a d-anti-s combination; that is, the K^0 meson. Each row of these diagrams corresponds to a different quark-colour designation which are of course indistinguishable, in accordance with the basic principle of the coloured quark system.



Accommodation of electric *Accommodation of strong* *Accommodation of weak*

We see that it is the z factor that comes to our rescue in the case of the K^0 particle. For if z were not always -1 for the strange-charm generation due to parity violation, we would have +2w and 0 as the charge structures for the K^0. This would have made it possible to identify the particular colours belonging to particular quarks in a particular combination, in violation of the basic principle of the coloured quark system.

The *s* component is the only one which takes the same value in all baryons or combinations of three unit charges and it is the only one which only ever occurs in one unit of charge at any given time. The 'colour' invariance requires that the single *s* component in the combination cannot be specified as belonging to one of the unit charges and a mechanism (which may be described as the strong interaction) must exist for a continual exchange of the *s* component between the three quarks in a baryon. Gauge invariance further requires that even the notion of an instantaneous location for *s* must be impossible. According to the tables in Appendix 2, transitions between systems A, B and C are equivalent to a transfer of the unit *s* component between the B, R and G quarks, without any change in the values of the *e* and *w* components, and the *s* component is the only one that can be transferred in this way. Consequently, it is the effective exchange of a single strong charge between the three bound quarks which prevents the identification of any one quark by its colour.

Transitions from systems A, B, C to systems D and E cannot be involved in the pure strong interaction because they involve the additional exchange of a weak component of charge.

The strong interaction has been accommodated via quantum field theory into a colour SU(3) group. An exact SU(3) symmetry is a precise equivalent of a continuous gauge-invariant transformation between three colour states as required by the simultaneous and equally probable existence of the three quark systems A, B and C.

The weak interaction may be found to occur because the D and E representations of quarks (see Appendix 2), cannot be obtained in general from any of the others without an exchange in weak, as well as strong, charges between the units. Transitions between C and D representations, in particular, or between B and E, require exchange of weak charges only and may be described as pure weak interactions.

Some decays of baryons and mesons provide evidence for the existence of another class of particles called the leptons. Just as in the case of quarks, the charge accommodation rules for leptons have an algebraic representation which explicitly describes three generations of leptons corresponding to the three generations of quarks. The weak interaction provides a mechanism for the transition between the A, B, C representations and the D representation. Since leptons are suspected to be the products of weak interactions between these representations, one would expect their structures to be evident within D. An inspection of red u, d, c, s quarks in the D system yields charge structures in weak and electric units identical to ν_e, e, ν_μ, μ and the γ.

Particle	Helicity	Charge Structure
e^-	left-handed	$-j\,e + k\,w$
e^-	right-handed	$-j\,e + 0\,k\,w$
ν_e	left-handed	$k\,w$
μ^-	left-handed	$-j\,e - z\,k\,w$
μ^-	right-handed	$-j\,e - 0\,z\,k\,w$
ν_μ	left-handed	$-z\,k\,w$
γ		$0\,j\,e + 0\,k\,w$

z in this case appertains to a violation of parity (z_P). The third generation of leptons (τ, ν_τ), appertains to z for violation of time reversal (z_T).

The leptons are unit charges with no s component. They behave in all respects like unbound quarks, and we may assume that it is the absence of the strong charge which makes them unbound. The mechanism for obscuring charge designations is not so strict now. Two charge types are now to be accommodated by allocating them to three quaternions.

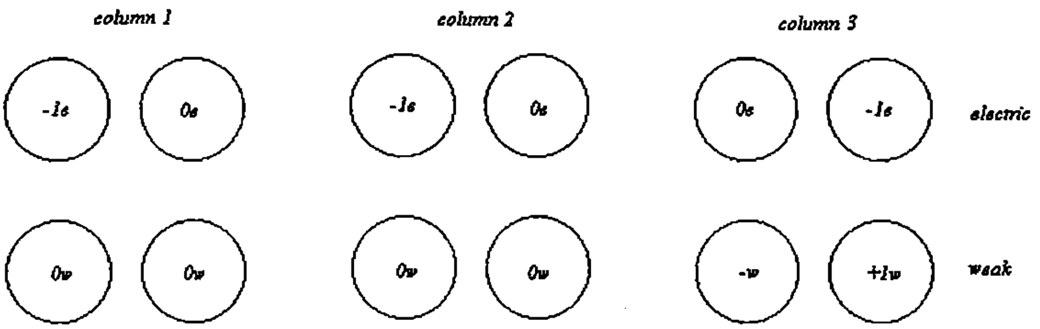
The leptons, however, are not completely free because weak interactions always require them to exist in a weak version of a bound state; that is leptons are always produced as part of a lepton-antilepton pair, like the quark-antiquark pairing of the meson. Though there is no strong binding involved as there is with quark-antiquark combinations, because of the complete removal of the s colour invariant system appropriate for three-component charges, the w and e charge components of leptons are still associated with quaternion operators which must remain unspecified.

We can see that the representations for the three generations of quarks collapses to three generations of leptons when there is no strong charge to accommodate. These generations still correspond to the three ways of accommodating the weak charge w; that is, $+w$ is indistinguishable from $-zw$, where z takes on the value of -1 for two reasons: 1) P is violated T is not and 2) T is violated and P is not. The table and matrix representation for leptons are identical to the tables and matrices for quarks with the omission of strong charge. (See Appendix 2.)

The representation above appertains of course to the D representation alone, because we have assumed that the lepton is the product of a weak transition between the A, B, C representations and the D representation. One should notice however, that in the absence of strong charge, A, B and C collapse to the same expression, which is different to D (and E). The physical significance of this will be investigated in our later work.

The construction of lepton charge structure is best demonstrated by appealing to an example. Let us construct the D representation of the down anti-up combination in the absence of strong charge:

Once again, the global charge structures of the various combinations are indistinguishable, and therefore charge allocation to quaternions has once again been obscured. However, the different possibilities alluded to by the diagrams must all be present for the full description of the down anti-up combination in the absence of strong charge. We see in these diagrams, familiar structures forming the couplets of lepton-antilepton pairs.



Various charge structures may be identified:
Column 1 represents the right handed electron (no weak charge) paired with a gamma.
Column 2 represents the right handed electron paired with a gamma
Column 3 represents the left handed electron paired with a right handed anti-electron neutrino.

6 The Standard Model

The rules for the Standard Model follow immediately from those for charge accommodation. The strong interaction follows an SU(3) symmetry whose exactness means that the exchange bosons or gluons remain massless. All possible 'colour' states in this representation are automatically identical. The electromagnetic charge structures for the quarks in the Standard Model (e.g., $2e/3$ for u, $-e/3$ for d, averaging out the values for the R, G and B representations) follow immediately from this colour-symmetry and its intrinsic gauge invariance.[15] With coupling constant g, generators of SU(3) $T^{\mu}/\,2$, and gauge fields $\mathbf{W}^{\mu}$, we are led naturally to a covariant derivative such as:

$$D^{\mu} = \partial^{\mu} + igT^{\mu}.\mathbf{W}^{\mu} / 2 .$$

The electric and weak interactions are linked in the SU(2) × U(1) symmetry which acts vertically in the quark tables, as the SU(3) acts horizontally. Here, it is significant that perfect symmetry between s, e and w would be maintained if we could restrict each of these charge-types to one sign only (+ in the case of s and w, –, for purely historical reasons, in the case of e). In this case, the representations A, B, C, D would then allow an SU(3) symmetry to be applied to each of the three interactions: A, B, C for s; B, C, D for e; A, C, D for w. However, the symmetry is broken when the opposite signs of the three charge-types are introduced.

For strong charges, the SU(3) symmetry remains, because all the $-s$ components can be assigned to antiparticles, though this implies that there must be a set of particles with $s = 0$ (leptons). The electromagnetic charge accommodates the opposite sign (+) by, in effect, adding 1 unit (k) to *all possible quark states*. This can be interpreted as a U(1) transition, or the addition of a scalar boson. It introduces the concept of isospin (I), which groups all particles with identical values of s and w, and the behaviour of the electromagnetic charge can be interpreted as the breaking of isospin symmetry. The concept of hypercharge (Y) then follows as 2 × the average charge of an isospin multiplet.

More significant than these are *weak* isospin (t, with third component t_3), which is an SU(2) symmetry relating all particles with the same value of w but disregarding the value of s, and the corresponding weak hypercharge (y), which deals with the electromagnetic (U(1)) component, the relationship between these and the average charge of a weak isomultiplet (Q) being given by:

$$Q = t_3 + y / 2 .$$

The specification of the third component of weak isospin is necessary because only one quark carries the weak charge. With coupling constants g and g', Pauli matrices τ as generators of SU(2), and gauge fields $\mathbf{W}^\mu$ and B^μ for weak isospin and weak hypercharge, the covariant derivative becomes

$$D^\mu = \partial^\mu + ig(\tau / 2).\mathbf{W}^\mu + i(g' / 2)yB^\mu .$$

The SU(2) means that two particle states (with different states of electric charge) are available for every value of w. Also, because of the particular nature of electric charge acommodation, no composite particles depend on the value of e; there is no electromagnetic equivalent of 'colour'. Hence, the infinite range of the interaction and the U(1) symmetry with which it is associated.

The three generations of weak isospin exist only to accommodate changes of sign in the weak charge. From the point of view of the weak interaction, however, there is only one generation and physical mixing can be expected to occur between all three generations. The mixing ensures that all weak interactions violate parity or time-reversal symmetry. For two generations, this involves the Cabibbo angle and Cabibbo-rotated eigenquarks of weak isospin; for three generations, the mixing needs to incorporate the CP-violating phase, as originally predicted by Kobayashi and Maskawa.[16]

A significant case of mixing ocurs between the K^0 and $\overline{K}^0$ meson states, where there is mixing between the parity-violating K^0 state with $(1 + z)\, \mathbf{k}$ and $\overline{K}^0$ antistate with $-(1 + z)\, \mathbf{k}$, each of which are also mixed with states with $0\, \mathbf{k}$. The $\mathbf{k}$ sign reversals mean that the *combinations* of K^0 and $\overline{K}^0$ create states with different parities, but with charge conjugation preserved (as must occur for a boson state); hence CP must also be violated. For the same reason, it is apparent that a similar combination of parity and CP violation must occur for the equivalent bosons in the third generation.

Other effects associated with the weak charge are strangeness and charm. Strangeness may be defined as $(3n - 3) / 2$, where n is the average number of weak units of charge (or units of $\mathbf{k}$) in a particle. Charm may be defined in a similar way, but it is associated with a quark with $+e$ instead of $-e$, and is also given the opposite sign by convention. Both strangeness and charm are measures of the degree of parity violation necessary to convert $-\mathbf{k}$ to $\mathbf{k}$. Similar measures could be applied in the next generation to the degree of violation of time-reversal. As strangeness and charm are purely properties of the $\boldsymbol{k}$ component, they are necessarily conserved in strong and electromagnetic interactions.

A strangeness-conserving current may be defined as one which retains the sign of $\boldsymbol{k}$, and a strangeness-changing current one which reverses it. The decay of Λ to p + e^- + $\bar{\nu}$ (equivalent to s + w or $\boldsymbol{i}$ + $\boldsymbol{k}$) requires a strangeness-changing current for those states of Λ with $-\boldsymbol{k}$, rather than $\boldsymbol{k}$, and so is slower than the same decay of the neutron, which always has $\boldsymbol{k}$. Another effect is that Ω^- is the only member of the spin 3/2 decuplet which always decays via a weak interaction because it always has a $-\boldsymbol{k}$ weak component, unlike any other state in the octet or decuplet, and can only decay to a state which has some fraction of $+\boldsymbol{k}$.

The charge structures of neutrinos show, at least in the first two generations, the clear distinction between electron and muon neutrinos, which require, respectively, $+w$ and $-w$ or $z_P w$. It is an interesting question whether $-w$ or $z_P w$ can be distinguished physically from the $-w$ or $z_T w$, representing the tau neutrino, or whether they are mixed or oscillating states.

With the SU(3) and SU(2) × U(1) symmetries established from fundamental principles, it is possible to use the Lagrangians already available for QCD and the Weinberg-Salam model, and justify them term by term. (This will be the subject of a later paper.) However, new numerical information can be extracted once the Grand Unified gauge group is identified, while particle masses can be assumed to be the result of the breaking of various symmetry conditions (in particular those involved with the weak interaction) and the suppression of units of each of the charges.

Baryon and lepton conservation are obvious consequences of the separate conservation laws for s and w. Baryons are the only particles containing both s and w, and so there is no decay mode available to eliminate the strong charge. Leptons, similarly, are the only particles with w, but no s, and so there is again no decay mode available to create another state.

The charge-accommodation rules also clearly point to the distinction between fermions and bosons. To describe the full range of fermions, we require a charge quaternion structure, whereas bosons require only the specification of the electromagnetic charge (and then only in the form of the added scalar term). The charge quaternion structure for fermions leads on to a requirement for a mass-quaternion wavefunction and Fermi-Dirac statistics, in contrast to the scalar wavefunction and Bose-Einstein statistics required of bosons. In particular, fermions have $+\boldsymbol{k}$ (either directly or through parity or time-reversal violation); bosons have $0\boldsymbol{k}$. The SU(2) spin ± 1/2 symmetry maps directly onto the identical symmetry for weak isospin.

The derivation of the spin 1/2 value from the $+\boldsymbol{k}$ component also means that the proton spin must be treated as a global feature and not derived from three component ± 1/2 spins from the quarks.

7 Particle Masses

If we take the rest mass of the electron as fundamental and determined in some way by the value of the electromagnetic charge, we may suppose that masses which originate in the strong coupling between quarks should take values related to the electron's mass by the term $1 / \alpha$, which is the ratio of the strong to electromagnetic couplings.

For energy balance between the different degrees of colour freedom of a unit charge, it is reasonable to suppose that each component of charge is equivalent to $m_e c^2 / \alpha$. Where charge components are absent the energy balance will require the existence of a mass of m_e / α, or kinetic energy equivalent. Since the allocation of zero and unit charges within quarks is a product of the process of quark confinement, we may suppose that energy which results in some way from the process must itself be energy of confinement. We could therefore imagine a vacuum with non-zero zero-point enery from which excitations could occur leading to components of charge equivalent to $m_e c^2 / \alpha$.

The Higgs mechanism attaches mass values to fermions by coupling them with the Higgs field, which represents a vacuum state with non-zero mass; and it is the non-zero vacuum state which is held responsible for symmetry-breaking. In the system outlined here, fermions gain mass when they lose charge, and the absence of charge components is the mechanism which breaks the symmetry. This means that zero charge is associated with finite mass, the charge component being effectively replaced by the mass unit.[17] However, total zero charge would be vacuum, and would also be equivalent to mass. Such zeroing would presumably only be possible in a boson state, where – in contrast to fermion states – the global charge structure is allowed to be zero. We could suppose, then, that the Higgs particle is equivalent to a boson with 'quark-antiquark' pairing, but with completely zero-charge in its components. Its coupling could, therefore, be considered equivalent to supplying particles with zero charges, and it would also be a vacuum state with nonzero mass.

Any attempt to obtain the mass of the Higgs boson must be, at this stage, highly speculative, since the details of the mechanism have still to be worked out. We may begin, however, by supposing that the Higgs represents the zeroing of all charge states under all possible representations. Taking 6 flavours, 6 antiflavours, 3 colours, 3 charges for each quark / antiquark, and 2 for each quark / antiquark pairing, over 4 representations, gives a total of 2596 zeros, which, *on the assumption that* each represents a mass-energy of $m_e c^2 / \alpha$, gives an approximate total of 182 GeV. This increases to 3240 $m_e c^2 / \alpha$ or 227 GeV if the fifth representation is included.

In considering the masses associated with various baryons and mesons, we need to consider the symmetries such as $SU(3)_f$ which group together particles in various isospin multiplets. With the electromagnetic charge component within any given isopsin multiplet being determined purely by a scalar addition term, and with the strong interaction being electromagnetic charge independent, we consider such multiplets to represent a single particle in different states of electromagnetic charge. All states of one multiplicity exist simultaneously, and so the zero charge components of all states must be accommodated in determining the mass of the particle.

The global SU(3)$_f$ symmetry shows the spin 3/2 baryons as a decuplet, with four Δ states, three Σ states, two Ξ states and one Ω state. The zero charge components of the Δ particles are simply those of all four states added together, but the four Δ states, when excited. have to be averaged between three Σ states, so each Σ state represents an average of 4/3 states. and the average number of charge components has to be multiplied by 4/3. The four Δ states are eventually ecited to one Ω state. so each Ω state represents an average 4 states. If M_0 is the highest multiplicity in a particular baryon octet or decuplet, and n_0 is the total number of zero charges in the components of a multiplet of multiplicity M, then the minimum mass of the components of the multiplet is given by

$$\text{mass} = \frac{n_0 M_0}{M} \frac{m_e}{\alpha} .$$

For example, the multiplet Σ in the spin 3/2 baryon decuplet has M = 3, and M_0 = 4 (the multiplicity of the Δ particles); while n_0, the total number of zeros, computed from the quark tables and circle diagrams, in the combinations dds, uds and uus, is 15, 17 or 19. For the ground state,

$$\text{mass of } \Sigma \text{ multiplet} = \frac{15 \times 4}{3} \frac{m_e}{\alpha} = 20 \frac{m_e}{\alpha} ,$$

which may be compared wth the experimental value of 19.8 m_e / α or 1385 MeV.

The derivation of the masses of the spin 3/2 decuplet may be set out in the following table:

	quark structure	n_0	M_0	M	predicted mass	measured mass
Δ	ddd,udd,uud,uuu	20,22,24	4	4	20 m_e / α	$\approx$ 17.6 - 19.6 m_e / α
Σ	dds,uds,uus	15,17,19	4	3	20	19.8
Ξ	dss.uss	11,13	4	2	22	21.9
Ω	sss	6	4	1	24	23.9

The masses are all calculated using the ground state values for n_0. The Δ particle is unusual in showing a large spread of measured mass; this is because, in this case, the energy width (approaching 2 m_e / α at half-maximum) makes a significant contribution, in addition to the rest mass; this energy width is much greater than that for any other member of the decuplet and explains the particle's instability and very rapid decay. The rest mass value (17.6 m_e / α) preserves the difference of 2 m_e / α between each multiplet which occurs due to successive transitions of one d quark to s with the net loss of two w charges. (2ijm is, notably, the term added on to the wavefunction in the violation of charge conjugation in the weak interaction.) The increasing accuracy of the predictions. from Δ through to Ω, may be related to the fact that the heavier particles represent fewer alternative states.

A similar analysis may be applied to the spin 1/2 baryon octet, though here the value for n_0 is taken at the ground state for the N multiplet (n, p), which contains no s quark component, and the other mass values are assumed to be of mixed states determined from within the predicted range by their accommodation within the Gell-Mann-Okubo formula required for $SU(3)_f$:

$$\frac{1}{2}(N + \Xi) = \frac{3}{4}\Lambda + \frac{1}{4}\Sigma$$

(where the particle symbols are used for convenience to represent the masses).

	quark structure	n_0	M_0	M	predicted mass	measured mass
N	udd,uud	9,11,13	3	2	13.5 m_e / α	13.4 m_e / α
Λ	uds	5,7	3	1	15 - 21	15.9
Σ	dds,uds,uus	15,17,19	3	3	15 - 19	17
Ξ	dss,uss	11,13	3	2	16.5 - 19.5	18.9

It is noticeable once again that the splitting of Σ and Ξ is almost exactly 2 m_e / α.

The meson octets do not represent the regular progression of excited states from the lowest member which we observe in the baryon octet and decuplet, and which ultimately derive from d → s quark transitions. The multiplets are, in this sense, independent, with mass determined by $n_0 \, m_e / \alpha$, where n_0 is the number of zero charge components in the multiplet. For the pseudoscalar 0^- meson octet, the ground state value of n_0 is once again chosen for the lowest lying member of the octet (π) – which again contains no symmetry-breaking s quark component – and the values for K and η selected from within the predicted range to fit a Gell-Mann-Okubo formula for $SU(3)_f$:

$$K^2 = \frac{1}{4}\pi^2 + \frac{3}{4}\eta^2 .$$

n_0 is 2, 6, 8, 10, 12, 14, or 16 for π, so the ground state value is 2; the predicted mass is therefore 2 m_e / α, which is exactly the observed value. For K, n_0 takes values 3, 5, 7, 9 or 11, and so the predicted mass is between 3 and 11 m_e / α, compared with the observed mass of 7.1 m_e / α; while η (which is additionally mixed with a singlet state) has n_0 values of 4, 6, 8, 10, 12, leading to a predicted mass between 4 and 12 m_e / α, compared with an observed mass of 7.8 m_e / α.

8 The Charge-Mass Mapping

As previously stated, the Dirac equation is no more than a quantized version of the classical relativistic conservation of energy equation:

$$E^2 - p^2 = m^2 ,$$

with the operator

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm \right)$$

acting on the wavefunction

$$\psi = (kE + ii\,\mathbf{p} + ijm)\, e^{-i(Et - \mathbf{p.r})}$$

producing the expression

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)\psi = (kE + ii\,\mathbf{p} + ijm)(kE + ii\,\mathbf{p} + ijm)\,e^{-i(Et - \mathbf{p.r})} = 0\,.$$

In the classical version, E and p are variables within the restriction that $E^2 - p^2$ is a constant for fixed m^2. However, for stationary quantum states, E and p are fixed, along with m, and the variability is confined to the space and time parameters in the exponential term of the wavefunction.

In classical relativistic theory, it is usual to emphasize the 4-vector nature of (iE, $\mathbf{p}$) and to describe $E^2 - p^2$ as an invariant, but, to incorporate this invariance directly, we need to define a new term with five components. Hence the expression ($kE + iip + ijm$) used in the Dirac wavefunction. Here we are using a 5-'dimensional' quantity to combine the effects of 3-dimensional conserved and nonconserved parameters (the term p having 3 dimensions, although only one is normally defined). A term of this kind might be conveniently referred to as a '5-vector' or 'Dirac operator'.

In quantizing the energy-momentum conservation, in fact, the Dirac equation effectively structures mass (or energy-momentum-mass) as a 3-dimensional conserved parameter, like charge (with one of the 'dimensions' being itself dimensional). We might therefore expect the same logic to apply to '3-dimensional' conserved mass as it does to 3-dimensional conserved charge. A one-to-one mapping of charge onto mass might be of the form:

E	p	m
w	s	e

The p / s relationship is immediately apparent, since both are 3-'dimensional' quantities (with only one dimension normally defined). The weak charge may be thought to have characteristics like those of the energy parameter, since the latter is the imaginary part of a 4-vector, with ± components which are ambiguous in relation to those of p and m. (The related time variable also has the property of being physically unidirectional while possessing both positive and negative directions of mathematical symmetry.) Again, mass units are like those of e in being added / subtracted only in fixed amounts. The pseudoscalar iE thus presents a natural SU(2) just as the vector $\mathbf{p}$ presents a natural SU(3) and the scalar m a natural U(1). Further evidence for this mapping may be found in the fact that the ratios w / s and E / p uniquely determine the spin-values for stationary states. (A similar mapping may exist for the 5 representations: A, B, C certainly represent the strong interaction, with D and E seemingly involved in the electroweak interaction.)

Since k is the component responsible for the SU(2) weak isospin symmetry, with eigenstates ± 1/2, it is not surprising to find (as noted above) that it has an exact analogy with the SU(2) spin ± 1/2 states of fermions derived from the Dirac equation. This leads on to the question of whether the wavefunction components can be mapped directly onto the quaternion charge operators describing particle structures, as suggested by the matrix representation. It may be significant that the absence of k and i reduces the boson charge structure to a term in j, which is equivalent to the scalar term m in the Dirac wavefunction. (A direct link is made when the combined signs of E and p are used to transfer between particle and antiparticle states.)

In principle, the Dirac equation introduces quaternion operators for E, p, m because of the need to remove cross-terms in the combination of squared quantities (in the same way as the individual conservation laws for e, s, w remove the cross-terms when these quantities are squared). In each case, the application of rotation symmetric quaternion operators to rotation asymmetric fixed quantities requires the application of the same symmetry-breaking (chiral) mathematical structure.

Further information on the charge-mass mapping may be derived from a consideration of the more fundamental symmetries relating space, time, mass and charge, outlined in section 3. The most significant symmetry here is that between continuous and discrete quantities. The 32-part algebra which underlies all physics based on the four parameters space, time, mass and charge has, as we have seen, a basis in five orthogonal states or primitive components, and there is a fundamental duality in applying this construction to the creation of particle states. If we apply it to the continuous quantity mass (represented in relativistic algebra by the energy-momentum-mass combination), we generate a Schrödinger-Dirac picture, characterised by wavefunctions, and with the five components representing energy, the three dimensions of momentum, and mass. If, one the other hand, we apply it to the discrete quantity charge, we generate a Heisenberg picture, with the five components representing the charges w, s (in three colour states) and e, and the five different ways of generating the appropriate 1 and 0 terms using random unit vectors in the charge-allocation representations A, B, C, D, E.

The random unit vectors produce zeros as well as units; so there must be five ways of producing zeros as well as unit charge components. Information about the 5-foldedness of the representation must come into the wavefunction that is created by the elimination of unit charge components by coupling to the Higgs field or vacuum, in the same way as it comes into the five independent ways of representing the fermion charge states. The pure energy state described as vacuum is suppressed in the production of discrete charge (removing the E, $\mathbf{p}$, m wavefunction representation); while the absence or suppression of charge and consequent breaking of the symmetry between the e, s and w interactions, creates the wavefunction components E, $\mathbf{p}$, m by the reverse process.

9 SU(5) Symmetry

The 5-fold nature of both the charge-accommodation system and the Dirac algebra suggests a possible relationship with an SU(5) symmetry for individual generations, as originally proposed by Georgi and Glashow.[18] Here, the 15 left-handed fermions are acommodated into SU(5) as follows:

$$\bar{5} = (\bar{3}, 1) + (1, 2)$$
$$= (\bar{d}) + (\nu_e, e^-)$$

$$10 = (\bar{3}, 1) + (1, 1) + (3, 2)$$
$$= (\bar{u}) + (\bar{e}^+) + (u, d)$$

It is significant here that the matrix representations of the charge-accommodation structures have terms which seemingly represent left-handed anti-electrons and right-handed electrons. Thus the D representation, whose R column represents the six left-handed leptons, also has terms whose charge structure adds up to *j*, exactly that expected of a left-handed anti-electron or anti-muon.

The gauge bosons, in SU(5), are obtained from

$$5 \times \bar{5} = 24 + 1$$
$$24 = (8, 1) + (1, 3) + (1, 1) + (3, 2) + (\bar{3}, 2)$$

The first term represents the 8 gluons, essentially the 8 non-singlet combinations of s and $\bar{s}$ states (the ninth being a colour singlet). The next two terms represent the four exchange particles for the combined electroweak interaction, W^-, W^+, Z^0 and γ, which are essentially the four possible combinations of (e, w) and $(\bar{e}, \bar{w})$. The last two terms are the six possible combinatons of s and $(\bar{e}, \bar{w})$. and the six possible combinations of $\bar{s}$ and (e, w), which are considered to be the X-bosons, mediating the combined strong-electroweak interaction. Significantly, the weak bosons W^-, W^+, and Z^0 and the X-particles are the only gauge bosons which acquire mass, because the weak interaction is the only one which involves changes in the sign of charge.

Exactly the same number of options are available to the terms in the Dirac wavefunction, e, s and w being now replaced by the respective terms m, p and E, and the respective combinations in the $5 \times \bar{5}$ by $p\ \bar{p}$, (m, E) $(\bar{m}, \bar{E})$, p $(\bar{m}, \bar{E})$, and $\bar{p}$ (m, E), or the equivalent vector / quaternion operators. The meaning of these terms may be taken as transitions between states represented by different wave functions, using annihilation and creation operators, the annihilation of one state and the creation of another. A transition like $p_y \rightarrow p_y$ would thus require a gauge boson like the colour transition R → G, or the representation transition A → E.

It has been considered that the existence of X-particles and a combined *s-e-w* interaction should necessarily lead to the decay of the proton via some such process as

$$\text{p} \rightarrow \pi^0 + \bar{e}^+ ,$$

in which a global charge structure $e + s + w$ is somehow reduced to the equivalent of $e - w$. However, such decays are forbidden by rotation asymmetry, and they also neglect the fact that weak interactions are subject to the requirements of what may (for want of a better term) be called 'weak colour' (the necessary creation of fermion-antifermion states, or their equivalent). It may be significant, however, that the ordinary process of positive beta decay can be described as a combined strong-electroweak interaction, without requiring either a violation of rotation asymmetry or of 'weak colour'.

The identification of the Grand Unified gauge group for one generation as SU(5) allows some numerical predictions to be made. Thus, the third component of weak isospin is ± 1/2 for left-handed states and 0 for right-handed. So

$$\Sigma\, t_3^{\,2}{}_{(u+d+e+\nu)} = 2\ .$$

Both right-handed and left-handed states, however, contribute to the electromagnetic charge, and

$$\Sigma\, Q_i^{\,2}{}_{(u+d+e+\nu)} = 8\ ,$$

rather than 16/3, as previously thought. This allows a prediction for the Weinberg angle from

$$\sin^2\theta_W = \Sigma\, t_3^{\,2} / Q_i^{\,2} = 0.25\ ,$$

in good agreement with the experimental value (far from Grand Unification) of 0.23.

It is conceivable also that the mapping of w, s, e onto E, p, m may extend to a relationship between the coupling constants for the three *interactions*, though the *group* coupling constants (g_1, g_2, g_3) all become identical under grand unification. If the value of $\sin^2\theta_W$ fixes the relationship between weak and electromagnetic couplings (here represented by e and w) by

$$w^2 \sin^2\theta_W = e^2\ ,$$

with

$$g_2^{\,2} = w^2 \text{ and } g_1^{\,2} = e^2 / \sin^2\theta_W\ ,$$

then it is possible that the relationship of these with the strong coupling constant (here represented by s) is of the form

$$w^2 = e^2 + s^2\ ,$$

by analogy with the relationship between E, p and m, which implies that

$$s^2 = 3e^2\ ,$$

or that

$$s^2 = w^2 \cos^2\theta_W\ ,$$

with

$$g_3^{\,2} = s^2 / \cos^2\theta_W\ .$$

10 Conclusion

Many of the most significant facts in particle physics have been shown to follow logically from a representation of the rotation asymmetric (conserved) charges s, e, w, by the rotation symmetric system of quaternions $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$. The SU(3) and SU(2) × U(1) components are immediately apparent, as is the SU(5) nature of the Grand Unified gauge group for one generation. Parity and CP violation, three generations, colour, bosons and fermions, baryons and mesons, quarks and leptons, and even the Higgs boson, can all be seen as consequences of a logical structure that has proved susceptible to a rigorous algebraic representation, and that can be linked to a parallel structure determining the algebra of the Dirac equation. Preliminary numerical work has begun with the determination of the Weinberg angle and the spectrum of particle masses.

Notes and References

1 P. Rowlands, *The Fundamental Parameters of Physics*, PD Publications, Liverpool 1991.

2 P. Rowlands, 'A new formal structure for deriving a physical interpretation of relativity, *Proceedings of Conference on Physical Interpretations of Relativity Theory II*, British Society for Philosophy of Science, London, September 1990, 264-8.

3 D. Hestenes, *Space-time Algebra*, Gordon and Breach, New York 1966..

4 W. Gough, 'Mixing scalars and vectors - an elegant view of physics', *Eur. J. Phys.*, **11**, 326-33, 1990.

5 P. Rowlands, 'The physical consequences of a new version of the Dirac Equation', in G. Hunter, S. Jeffers and J.-P. Vigier (eds.), *Causality and Locality in Modern Physics*, Kluwer 1998, pp. 397-402.

6 P. Rowlands, 'A new algebra for relativistic quantum mechanics', *Proceedings of Conference on Physical Interpretations of Relativity Theory V*, British Society for Philosophy of Science, London, September 1996, 381-7.

7 P. Rowlands, 'Some interpretations of the Dirac algebra', *Speculat. Sci. Tech.*, **19**, 243-51, 1996.

8 P. Rowlands and J. P. Cullerne, 'A symmetry principle for deriving particle structures', *Proceedings of Conference on Physical Interpretations of Relativity Theory VI*, British Society for Philosophy of Science, London, September 1998, 316-33.

9 P. Rowlands, 'Further considerations of the Dirac algebra' *Proceedings of Conference on Physical Interpretations of Relativity Theory VI*, British Society for Philosophy of Science, London, September 1996, 308-15.

10 I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics*, second edition, Adam Hilger, 1989.

11 Lewis H. Ryder, *Quantum Field Theory*, second edition, Cambridge University Press, 1996.

12 Francis Halzen and Alan D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, John Wiley, 1984.

13 P. Rowlands, 'Quantum uncertainty, wave-particle duality, and fundamental symmetries', in S. Jeffers, S. Roy, J.-P. Vigier and G. Hunter (eds.), *Causality and Locality in Modern Physics*, Kluwer 1998, pp. 361-72.

14 P. Rowlands, 'Quantum uncertainty, wave-particle duality, and the physical interpretation of relativity from first principles' *Proceedings of Conference on Physical Interpretations of Relativity Theory III*, British Society for Philosophy of Science, London, September 1992, 296-310.

15 See, for example, F. E. Close, *An Introduction to Quarks and Partons*, Academic Press, 1979, p. 233: 'if only colour singlet states are being produced then all transitions are from colour singlet to colour singlet and so the operative charges will be the $SU(3)_c$ average charges'.

16 M. Kobayashi and T. Maskawa, *Prog. Theor. Phys.*, **49**, 282, 1972.

17 That mass and charge are, in some sense, mutually exclusive components of the vacuum, is implied by standard treatments of the U(1) component of the Weinberg-Salam theory. See, for example, Aitchison and Hey, writing on the hypercharge value of the Higgs field: 'we do not allow the particle physics vacuum to give an electrically charged field a non-zero value. Thus we require that the component of ϕ with non-zero vacuum value has zero charge.' (*op. cit.* (ref. 9), p. 434..)

18 H. Georgi and S. L. Glashow, 'Unity of All Elementary-Particle Forces', *Phys. Rev. Letters*, **32**, 438-41, 1974.

Appendix 1

Units of the Dirac algebra

$1, \gamma^0 = ik, \gamma^1 = i\mathbf{i}, \gamma^2 = i\mathbf{j}, \gamma^3 = i\mathbf{k}, \gamma^5 = ij, \gamma^0\gamma^1 = ij\mathbf{i}, \gamma^0\gamma^2 = ij\mathbf{j}, \gamma^0\gamma^3 = ij\mathbf{k}, \gamma^0\gamma^5 = i, \gamma^1\gamma^2 = -ik\mathbf{k}, \gamma^1\gamma^3 = i\mathbf{j}, \gamma^1\gamma^5 = ik\mathbf{i}, \gamma^2\gamma^3 = -i\mathbf{i}, \gamma^2\gamma^5 = ik\mathbf{j}, \gamma^3\gamma^5 = ik\mathbf{k}, \gamma^0\gamma^1\gamma^2 = k\mathbf{k}, \gamma^0\gamma^1\gamma^3 = -k\mathbf{j}, \gamma^0\gamma^1\gamma^5 = \mathbf{i}, \gamma^0\gamma^2\gamma^3 = k\mathbf{i}, \gamma^0\gamma^2\gamma^5 = \mathbf{j}, \gamma^0\gamma^3\gamma^5 = \mathbf{k}, \gamma^1\gamma^2\gamma^3 = -ii, \gamma^1\gamma^2\gamma^5 = j\mathbf{k}, \gamma^1\gamma^3\gamma^5 = -j\mathbf{j}, \gamma^2\gamma^3\gamma^5 = j\mathbf{i}, \gamma^0\gamma^1\gamma^2\gamma^3 = j, \gamma^0\gamma^1\gamma^2\gamma^5 = -ii\mathbf{k}, \gamma^0\gamma^1\gamma^3\gamma^5 = ii\mathbf{j}, \gamma^0\gamma^2\gamma^3\gamma^3 = -ii\mathbf{i}, \gamma^1\gamma^2\gamma^3\gamma^5 = k, \gamma^0\gamma^1\gamma^2\gamma^3\gamma^5 = -i.$

Terms of the opposite sign are obtained by reversing the order of multiplication.

Appendix 2

A, B, C, D and E Representations

A

		B	G	R
u	$+e$	$1j$	$1j$	$0i$
	$+s$	$1i$	$0k$	$0j$
	$+w$	$1k$	$0i$	$0k$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$1i$	$0i$	$0k$
	$+w$	$1k$	$0j$	$0i$
c	$+e$	$1j$	$1j$	$0i$
	$+s$	$1i$	$0k$	$0j$
	$-w$	z_Pk	$0i$	$0k$
s	$-e$	$0i$	$0k$	$1j$
	$+s$	$1i$	$0i$	$0k$
	$-w$	z_Pk	$0j$	$0i$
t	$+e$	$1j$	$1j$	$0i$
	$+s$	$1i$	$0k$	$0j$
	$-w$	z_Tk	$0i$	$0k$
b	$-e$	$0i$	$0k$	$1j$
	$+s$	$1i$	$0i$	$0k$
	$-w$	z_Tk	$0j$	$0i$

B

		B	G	R
u	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$0k$	$1i$
	$+w$	$1k$	$0i$	$0j$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0k$	$1i$
	$+w$	$1k$	$0i$	$0k$
c	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$0k$	$1i$
	$-w$	z_Pk	$0i$	$0j$
s	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0k$	$1i$
	$-w$	z_Pk	$0i$	$0k$
t	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$0k$	$1i$
	$-w$	z_Tk	$0i$	$0j$
b	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0k$	$1i$
	$-w$	z_Tk	$0i$	$0k$

C

		B	G	R
u	+ *e*	1*j*	1*j*	0*k*
	+ *s*	0*i*	1*i*	0*j*
	+ *w*	1*k*	0*k*	0*i*
d	− *e*	0*j*	0*k*	1*j*
	+ *s*	0*i*	1*i*	0*k*
	+ *w*	1*k*	0*j*	0*i*
c	+ *e*	1*j*	1*j*	0*k*
	+ *s*	0*i*	1*i*	0*j*
	− *w*	$z_P k$	0*k*	0*i*
s	− *e*	0*j*	0*k*	1*j*
	+ *s*	0*i*	1*i*	0*k*
	− *w*	$z_P k$	0*j*	0*i*
t	+ *e*	1*j*	1*j*	0*k*
	+ *s*	0*i*	1*i*	0*j*
	− *w*	$z_T k$	0*k*	0*i*
b	− *e*	0*j*	0*k*	1*j*
	+ *s*	0*i*	1*i*	0*k*
	− *w*	$z_T k$	0*j*	0*i*

D

		B	G	R
u	+ *e*	1*j*	1*j*	0*i*
	+ *s*	0*k*	1*i*	0*j*
	+ *w*	0*i*	0*k*	1*k*
d	− *e*	0*i*	0*k*	1*j*
	+ *s*	0*j*	1*i*	0*i*
	+ *w*	0*k*	0*j*	1*k*
c	+ *e*	1*j*	1*j*	0*i*
	+ *s*	0*k*	1*i*	0*j*
	− *w*	0*i*	0*k*	$z_P k$
s	− *e*	0*i*	0*k*	1*j*
	+ *s*	0*j*	1*i*	0*i*
	− *w*	0*k*	0*j*	$z_P k$
t	+ *e*	1*j*	1*j*	0*i*
	+ *s*	0*k*	1*i*	0*j*
	− *w*	0*i*	0*k*	$z_T k$
b	− *e*	0*i*	0*k*	1*j*
	+ *s*	0*j*	1*i*	0*i*
	− *w*	0*k*	0*j*	$z_T k$

E

		B	G	R
u	$+e$	$1j$	$1j$	$0i$
	$+s$	$0k$	$0j$	$1i$
	$+w$	$0i$	$0k$	$1k$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0i$	$1i$
	$+w$	$0k$	$0j$	$1k$
c	$+e$	$1j$	$1j$	$0i$
	$+s$	$0k$	$0j$	$1i$
	$-w$	$0i$	$0k$	$z_P k$
s	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0i$	$1i$
	$-w$	$0k$	$0j$	$z_P k$
t	$+e$	$1j$	$1j$	$0i$
	$+s$	$0k$	$0j$	$1i$
	$-w$	$0i$	$0k$	$z_T k$
b	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0i$	$1i$
	$-w$	$0k$	$0j$	$z_T k$

Appendix 3

Particle Charge Structures

meson octet (0^-)

particle	quarks	charge structures	average structure
π^-	$u\,\bar{d}$	$+e$	
π^+	$d\,\bar{u}$	$-e$	
π^0	$u\,\bar{u}$	0	
η	$d\,\bar{d}$	0	
K^-	$u\,\bar{s}$	$-e$ or $-e-(1+z)\,w$	$-e-(1+z)\,w/3$
K^+	$s\,\bar{u}$	$-e$ or $-e-(1+z)\,w$	$-e-(1+z)\,w/3$
K^0	$d\,\bar{s}$	0 or $+(1+z)\,w$	$+(1+z)\,w/3$
$\bar{K}$	$s\,\bar{d}$	0 or $-(1+z)\,w$	$-(1+z)\,w/3$

baryon octet ($1/2^+$)

particle	quarks	charge structures	average structure
n	udd	$+s+w$	$+s+w$
p	uud	$+e+s+w$	$+e+s+w$
Λ	uds	$+s\pm w$	$+s+w/3$
Σ^-	dds	$-e+s\pm w$	$-e+s+w/3$
Σ^0	uds	$+s\pm w$	$+s+w/3$
Σ^+	uus	$+e+s\pm w$	$+e+s+w/3$
Ξ^-	dss	$-e+s\pm w$	$-e+s-w/3$
Ξ^0	uss	$+s\pm w$	$+s-w/3$

baryon decuplet ($3/2^+$)

particle	quarks	charge structures	average structure
Δ	ddd	$-e+s+w$	$-e+s+w$
Δ	udd	$+s+w$	$+s+w$
Δ	uud	$+e+s+w$	$+e+s+w$
Δ	uuu	$+2e+s+w$	$+2e+s+w$
Σ^-	dds	$-e+s\pm w$	$-e+s+w/3$
Σ^0	uds	$+s\pm w$	$+s+w/3$
Σ^+	uus	$+e+s\pm w$	$+e+s+w/3$
Ξ^-	dss	$-e+s\pm w$	$-e+s-w/3$
Ξ^0	uss	$+s\pm w$	$+s-w/3$
Ω^-	sss	$-e+s-w$	$-e+s-w$

PROGRAM UNIVERSE and RECENT COSMOLOGICAL RESULTS *

H. Pierre Noyes
Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94309

Abstract

Recent improvements in astronomical observations lead to the conclusion that the Hubble constant lies between 60 and 80 Mpc km^{-1} sec^{-1} and the age of the universe between 11 and 14 Gigayears. Taken together with recent observations of distant type Ia supernovae and the cosmic background radiation, these limits allow a check of the consequences of predictions made a decade ago using program universe and the combinatorial hierarchy that the ratio of baryons to photons is $1/256^4$ and of dark to baryonic matter is 12.7. We find that the restrictions on the matter content of the universe and the cosmological constant are within, and much tighter than, the limits established by conventional means. The situation is further improved if we invoke an estimate of the normalized cosmological constant made by E.D. Jones of $\Omega_\Lambda \sim 0.6$. This opens a "window of opportunity" to get the predictions of the ANPA program in front of the relevant professional community *before* precise observations lead to a consensus. We urge ANPA members to join us in the assault on this breach in the walls of establishment thinking.

*Work supported by Department of Energy contract DE–AC03–76SF00515.

1 Introduction

When Fredrick (Parker-Rhodes) discovered the combinatorial hierarchy [1] in 1961, the excitement arose from the successful calculation of two dimensionless, empirical ratios—the fine structure constant and the ratio of electromagnetic to gravitational forces. These numbers were already known to physicists to better accuracy than his calculation provided, but (then and now) no extant conventional theory provided a way to calculate them. These same facts held true for his subsequent calculation of the electron-proton mass ratio [2], and also for many numbers the ANPA program has produced over the years. I discussed some of the reasons why established physicists continue to ignore these results in my introductory lectures presented here a couple of years ago [3]. One basic reason is that the numbers were known before the calculations were made, leaving the program open to a charge of engaging in "numerology".

If I had been lucky, I might have been able to predict that there are only three generations of neutrinos before SLAC and LEP demonstrated this experimentally, but I doubt that this would have made much difference to the reception of our results by most physicists. The basic difficulty remains that our line of reasoning is so foreign to most physicists that any success we have along these lines will need to be (a) dramatic, (b) timely, and (c) well publicized in the relevant professional literature *before* the observations are made. It looks unlikely that these conditions will be met any time before ANPA 40, so far as particle physics goes.

One reason for this paper is to point out that we may now have a better chance of getting our cosmological predictions before the relevant audience in a timely fashion than we have had with particle physics. But this "window of opportunity" may easily slip by us unless more effort is put on cosmological predictions than we have exerted in the past. Here-to-fore the basic observational cosmological parameters have been so uncertain, and the competing "conventional" theories so multifarious and speculative, that our rather precise results (for those of us who believe in *program universe* [4]) seemed to have little prospect of getting attention, let alone confirmation. This situation has changed quite dramatically in the last year thanks to a number of different results that restrict the Hubble constant to the range 60 to 80 km sec^{-1} Mpc^{-1} [5, 6], limit the age of the universe to between eleven and fourteen Gigayears

[5], require the universe to have much (and perhaps more) of its expansion rate determined by a repulsive cosmological constant rather than by the lack of closure mass, and give direct measurements [7] of dark matter between the galactic clusters as well as of the dark matter surrounding them.

The almost totally unexpected result that the universe is *demonstrably* going to keep on expanding forever was implied by the physical interpretation of *program universe* tabulated in [8] a decade ago. What I called M_{vis} in that table was meant to imply "electromagnetically observable" rather than "detected by recording visible light". This ambiguity persists in the table published [9] in 1994, but the prediction presented there of the number ratio of baryons to photons, first published [10] in 1991, correctly recognizes that *program universe* gives, as a first approximation, baryonic matter rather than the observational parameter sometimes called "visible matter".

The distinction between baryonic and observationally visible matter is significant because a possibly substantial and currently unknown fraction of the baryonic matter may occur in the form of "brown dwarfs". What is important here is that our theory gives what turns out to be quite a good prediction of the baryon to photon ratio *and* of the ratio of dark to baryonic matter. The energy density in photons is known directly from the temperature of the cosmic background radiation. Further, since our prediction only has two categories of matter in significant quantities in this epoch, we get a prediction of the *total* matter density independent of answering the vexed question of how much of the non-visible but baryonic matter is in the form of brown dwarfs or other ordinary matter. Thus, knowing the Hubble constant H_0, we can predict $\Omega_M \equiv \rho_m/\rho_c$ where $\rho_c = 3H_0^2/8\pi G_N$, G_N being the Newtonian gravitational constant.. The absolute mass of the universe we also predicted a decade ago is still not easily connected to observational data, but even in 1989 pointed in the direction of an open universe [8], in opposition to a near consensus among the cosmological theorists.

As we will show in Section 3, taking seriously our predicted ratio of baryons to photons and of dark to baryonic matter is already quite restrictive, and well within the limits allowed by current cosmological observations. When augmented by an "a priori" estimate of the cosmological constant made by Ed Jones (cf. Section 4), we end up being able to make a prediction of the two parameters (Ω_M and Ω_Λ) which specify the gross cosmology of the universe that is *better* than current observations can test. The three numbers M_{Dark}/M_B, N_B/N_γ and $M_{\rm Univ}$ play a role in observational cosmology comparable to the role played by the three Parker-Rhodes numbers $(e^2/\hbar c, m_{\rm Planck}/m_{\rm proton}, m_p/m_e)$ in particle physics. In both observational cosmology and elementary particle physics the conventional approach requires the numbers to be taken from observation and fitted into an hypothesized theory rather than calculated from first principles. In contrast to the empirically well known Parker-Rhodes numbers, however, the first two cosmological numbers are only this year beginning to take on consensus values at the ten to twenty per cent level, and $M_{\rm Univ}$ is still know only within a factor of two or so. This time the ANPA program has a fighting chance to get our numbers in front of the relevant audience *before* they are well measured rather than *after* the fact.

How the "repulsive cosmological constant" comes about is another story, which will only be briefly touched on in this paper. Ed Jones had reached that result on quite general grounds some time ago, but unfortunately did not publish his conclusion because of lack of observational evidence. He is now preparing a short paper on the subject [11], which I pray will get into the literature in time to get him some credit.

But none of my remarks here will make sense unless we have in front of us the recent observational results which have so dramatically changed the cosmological picture.

2 A Brief Survey of Recent Cosmological Results

A number of factors, which are the culmination of many years of hard work by many astronomers, astrophysicists, and physical cosmologists, have converged rather suddenly on definite observational cosmological results. Partly these are simply the result of the accumulation of data from the large Keck telescopes in Hawaii and Chile, as well as from smaller observatories, and from the Hubble Space Telescope. Partly they are the result of accumulating satellite data, particularly from Hipparcos and the COsmic Background Experiment (COBE). But the data analysis would not have been possible without the increasing power and availability of low cost computers, and would not have yielded such dramatic results so quickly without some very clever ideas exploiting the technological and observational opportunities.

Before plunging into my description of some of the results, I wish to stress that I am an outsider in this field, and have had to rely almost exclusively on one conference this spring (DM98) and one summer institute (SSI XXVI) which I had the good fortune to attend. There was enough discussion of controversial matters in both of these environs to allow me to believe I could assess reasonably accurately the outlines of agreement that are emerging, and to draw my own conclusions. But be warned that my lack of background could have led me pretty far astray.

One basic fact which seems pretty firm is that Hipparcos has now supplied a large enough sample of cephid variables with measured parallax to change the calibration of the distance-luminosity relation by ten percent or more. So far as the age of the universe goes, it is equally important that the parallax of several globular clusters has also been measured. This turned out to make the oldest objects in our own galaxy younger than estimates of the "age of the universe" by the right amount to achieve consistency [12, 5]. Thus there is no longer an "age problem". The universe, and its contents, have existed something like 12.5 billion (1.25×10^{10}) years, give or take a billion or so; as our reference time we take the backward extrapolation to the time when the contents of the universe must have been so compacted that the question of whether we can trust the laws of physics enough to extrapolate any earlier becomes, for some of us, the critical question. Fortunately both "fireball time" (when the radiation breaks away from the matter) and the earlier time of nucleosynthesis (when the neutrons freeze out and for a few minutes can be used to form deuterons, ^3He, alpha particles and ^7Li) are sufficiently later than this "epistemological cutoff" so that we can still perform relevant laboratory experiments to check our assumptions. We can remain comfortable with the cosmological calculations needed in what follows from an operational point of view.

This extrapolation back in time starts from "local" evaluations of the "Hubble Constant"—the distance-velocity relation between recessional velocity as measured by red-shift using data out to about 100 Mega-parsec ($1\ Mpc = 3.26... \times 10^6$ light years). For the nearer galaxies this again depends on the recalibration of the cephids, but also on getting a handle on the local imhomogenieties (Virgo cluster, the "great attractor", etc., etc.). Recently much more data has become available on the "Peculiar Velocities" of Galaxies which deviate from the average Hubble streaming. Consequently one can plot the overall distribution of gravitating matter (rather than the distribution of light) over this enormous—but still "local"—region. These measures have to be self-consistent. When this is achieved, as is claimed, it reinforces the conclusion, which now comes from several different types of data, that most of the gravitating matter in the universe is dark rather than luminous.

To take the Hubble relation back farther, one needs a "standard candle" that is reliable to as early times as is possible. It turns out that the type Ia supernovae are numerous enough in the region where cephid measurements can still be made to collect enough calibrated light curves to establish what is needed. This took a very clever combination of physical reasoning and optimal utilization of resources which have to be shared with many other meritorious observational programs. It is this data which gives firm evidence for a repulsive cosmological constant [13].

Direct measurement of the dark matter itself has been made by detailed analysis of the defects in the gravitational lenses provided by clumps of "local" galaxies imaging very early galaxies (back to red-shift 5!) [7]. Most of the lensing comes from the dark matter itself, not from the sprinkling of visible matter which, presumably, has fallen into it. Such lenses also exist between visible clumps; these lenses may or may not include burned out galaxies, but are not optically visible. These results confirm the hypothesis that much more of the gravitating matter in the universe is dark than luminous. Even with this additional matter, there is not enough to close the universe in the absence of an attractive cosmological constant, let alone enough in the presence of the observed repulsive cosmological constant. This becomes clear when the COBE data and the type Ia supernovae data are combined [14, 15].

3 Consequences of Two Program Universe Predictions

3.1 Program Universe

Here we remind the reader of how we use *discrimination* ("$\oplus$") between ordered strings of zeros and ones (*bit-strings*) defined by

$$(\mathbf{a}(W) \oplus \mathbf{b}(W))_w = (a_w - b_w)^2; \quad a_w, b_w \in 0,1; \quad w \in 1,2,....,W \tag{1}$$

to generate a growing universe of bit-strings which at each step contains $P(S)$ strings of length S. We use an algorithm known as *program universe* which was developed in collaboration with M.J.Manthey [16, 4]. Since no one knows how to construct a "perfect" random number generator, we cannot in practice start from Manthey's "flipbit" (which returns a zero or a one with equal probability when asked), and must content ourselves with a pseudo-random number generator that, to some approximation which we will be wise to reconsider from time to time, will come close to that performance. Using any available approximation to "flipbit" and assigning an order parameter $i \in 1,2,...,P(S)$ to each string in our array, Manthey [16] has given the coding for constructing a routine "PICK" which picks out some arbitrary string $\mathbf{P}_i(S)$ with probability $1/P(S)$. Then program universe amounts to the following simple algorithm:

PICK any two strings $\mathbf{P}_i(S)$,$\mathbf{P}_j(S)$, $i,j \in 1,2,...,P$ and compare $\mathbf{P}_{ij} = \mathbf{P}_i \oplus \mathbf{P}_j$ with $\mathbf{0}(S)$.

If $\mathbf{P}_{ij} \neq \mathbf{0}$, adjoin $\mathbf{P}_{P+1} := \mathbf{P}_{ij}$ to the universe, set $P := P+1$ and recurse to PICK. [This process is referred to as ADJOIN.]

Else, for each $i \in 1,2,...,P$ pick an arbitrary bit $\mathbf{a}_i \in 0,1$, replace $\mathbf{P}_i(S+1) := \mathbf{P}_i(S) \| \mathbf{a}_i$, set $S := S+1$ and recurse to PICK. [This process is referred to as TICK.]

Here the operation "$\|$" simply extends the string on the left of the symbol by adjoining the string to its right (in the instance above, the arbitrary bit $\mathbf{a}_i$ supplied by "flipbit") and adjusting the ordering indices and resulting string length parameter appropriately. We note that any universe so generated is "uncrunchable", to quote John Wheeler [17]. In our current context this construction, taken seriously, *necessarily* requires that the cosmological constant be greater than zero, as we will assume below.

3.2 Events, Labels, Contents

This version of *program universe*—called "Program Universe 2" in the published Ref. [8]—provides considerable structure to "events", modeled by the two alternatives presented above. Note that so long as the string produced by the event is non-null (and hence that all three strings are non-null and different from each other), the string length does not change (i.e. there is no TICK). Interpreted as a three-leg Feynman diagram (a story we cannot develop to any great extent in this paper), ADJOIN can be shown to correspond to a "vacuum fluctuation" which conserves (relativistic) 3-momentum but not energy, and hence is unobservable as a physical process. On the other hand, when two indistinguishable strings are compared, producing a TICK, this can be interpreted as four-leg Feynman diagram in which one of the two indistinguishable strings was produced earlier and the other serves as the needed spectator in any observable *relativistic finite particle number* three body scattering process [18].

Program Universe 2 also provides a separation into a conserved set of "labels", and a growing set of "contents" which can be thought of as the space-time "addresses" to which these labels refer. To see this, think of all the left-hand, finite length S portions of the strings which exist when the program TICKs and the string-length goes from S to $S+1$. Call these *labels* of length $L = S$, and the number of them at the critical tick $N_0(L)$. Further PICKs and TICKs can only add to this set of labels those which can be produced from it by pairwise discrimination, with no impact from the (growing in length and number) set of content labels with length $S_C = S - L > 0$. If $N_I \leq N_0(S_L)$ of these labels are *discriminately independent*, then the maximum number of distinct labels they can generate, no matter how long program universe runs, will be $2^{N_I} - 1$, because this is the maximum number of ways we can choose combinations of N_I distinct things taking them $1, 2, ..., N_I$ times. We will interpret this fixed number of possibilities as a representation of the quantum numbers of systems of "elementary particles" allowed in our bit-string universe and use the growing content-strings to represent their (finite and discrete) locations in an expanding space-time description of the universe.

This label-content schema then allows us to interpret the events which lead to TICK as four-leg Feynman diagrams representing a stationary state scattering process. Note that for us to find out that the two strings found by PICK are the same, we must either pick the same string twice or at some previous step have produced (by discrimination) and adjoined the string which is now the same as the second one picked. Although it is not discussed in bit-string language, a little thought about the solution of a relativistic three body scattering problem Ed Jones and I have found [18] shows that the driving term ($\substack{> - < \\ -}$) is always a four-leg Feynman diagram ($> - <$) plus a spectator ($-$) whose quantum numbers are *identical* with the quantum numbers of the particle in the intermediate state connecting the two vertices. The step we do not take here is to show that the labels do indeed represent quantum number conservation and the contents a finite and discrete version of relativistic energy-momentum conservation. But we hope that enough has been said to show how we could interpret program universe as representing a sequence of contemporaneous scattering processes, and an algorithm which tells us how the space in which they occur expands..

3.3 Cosmological Interpretation of Program Universe

At this point we need a guiding principle to show us how we can "chunk" the growing information content provided by discriminate closure in such a way as to generate a hierarchical representation of the quantum numbers that the label-content schema provides. Following a suggestion of David McGoveran's [19], we note that *we can guarantee that the representation has a coordinate basis and supports linear operators by mapping it to square matrices.*

The mapping scheme originally used by Amson, Bastin, Kilmister and Parker-Rhodes [20] satisfies this requirement. This scheme requires us to introduce the multiplication operation ($0 \cdot 0 = 0 = 0 \cdot 1 = 0 = 1 \cdot 0$, $1 \cdot 1 = 1$), converting our bit-string formalism into the *field* Z_2. First note, as mentioned above, that any set of n discriminately independent (*d.i.*) strings will generate exactly $2^n - 1$ discriminately closed subsets (*dcss*). Start with two d.i. strings $\mathbf{a}$, $\mathbf{b}$. These generate three d.i. subsets, namely $\{\mathbf{a}\}$, $\{\mathbf{b}\}$, $\{\mathbf{a}, \mathbf{b}, \mathbf{a} \oplus \mathbf{b}\}$. Require each dcss ($\{\quad\}$) to contain only the eigenvector(s), of three 2×2 *mapping matrices* which (1) are non-singular (do not map onto zero) and (2) are d.i. Rearrange these as strings. They will then generate seven dcss. Map these by seven d.i. 4×4 matrices, which meet the same criteria (1) and (2) just given. Rearrange these as seven d.i. strings of length 16. These generates $127 = 2^7 - 1$ dcss. These can be mapped by 127 16×16 d.i. mapping matrices, which, rearranged as strings of length 256, generate $2^{127} - 1 \approx 1.7 \times 10^{38}$ dcss. But these cannot be mapped by 256×256 d.i. matrices because there are at most 256^2 such matrices and $256^2 \ll 2^{127} - 1$. Thus this *combinatorial hierarchy* terminates at the fourth level. The mapping matrices are not unique, but exist, as has been proved by direct construction and an abstract proof [21]. It is easy to see that the four level hierarchy constructed by these rules is *unique* because starting with d.i. strings of length 3 or 4 generates only two levels and the dcss generated by d.i. strings of length 5 or greater cannot be mapped.

Making physical sense out of these numbers is a long story [3], and making the case that they give us the quantum numbers of the standard model of quarks and leptons with exactly 3 generations has only been sketched [9]. However we do not require the completely worked out scheme to make interesting cosmological predictions. The ratio of dark to "visible" (i.e. electromagnetically interacting) matter is the easiest to see. The electromagnetic interaction first comes in when we have constructed the first three levels giving 3+7+127 =137 dcss, one of which is identified with electromagnetic interactions because it occurs with probability $1/137 \approx e^2/\hbar c$. But the construction must first complete the first two levels giving 3+7=10 dcss. Since the construction is "random" and this will happen many, many times as program universe grinds along, we will get the 10 non-electromagnetically interacting labels 127/10 times as often as we get the electromagnetically interacting labels. Our prediction of $M_{Dark}/M_B = 12.7$ is that naive. We discuss how we might improve the calculation of this number in the concluding section.

The $1/256^4$ prediction for N_B/N_γ is comparably naive. Our partially worked out scheme of relating bit-string events to particle physics [9, 3], makes it clear that photons, both as labels (which communicate with particle-antiparticle pairs) and as content strings will contain equal numbers of zeros and ones in appropriately specified portions of the strings. Consequently they can be readily identified as the most probable entities in any assemblage of strings generated by flipbit. This scheme also makes the simplest representation of fermions and anti-fermions contain one more "1" or one more "0" than the photons. (Which we call "fermions" and which "anti-fermions" is, to begin with, an arbitrary choice of nomenclature.) Since our dynamics insures conventional quantum number conservation by construction, the problem — as in conventional theories—is to show how program universe introduces a bias between "0" 's and "1" 's once the full interaction scheme is developed. (The recently commissioned "B-Factory at SLAC is aimed at providing experimental evidence relevant to a conventional explanation of the observed bias between matter and anti-matter in our universe.)

Since program universe has to start out with two strings, and both of these cannot be null if the evolution is lead anywhere, the first significant PICK and discrimination will necessarily lead to a universe with three strings, two of which are "1" and one of which is "0". Subsequent PICKs and TICKs are sufficiently "random" to insure that (at least statistically) there will be an equal number of zeros and ones, apart from the initial bias giving an extra one. Once the label length of 256 is reached, and sufficient space-time structure ("content strings") generated and interacted to achieve thermal equilibrium, this label bias for a 1 compared to equal numbers of zeros and ones will persist for 1 in 256 labels. But to count the equilibrium processes relevant to computing the ratio of baryons to photons, we must compare the labels leading to baryon-photon scattering compared to those leading to photon-photon scattering. This requires the baryon bias of 1 to appear in one and only one of the four labels of length 256 involved in that comparison; this comparison is illustrated in Fig. 1, which assumes that the above mentioned interpretation of the strings causing observable TICK's as four leg Feynman diagrams has been satisfactorily demonstrated. We conclude that, in the absence of further information, $1/256^4$ is the program universe prediction for the baryon-photon ratio at the time of big bang nucleosynthesis. We will discuss how this estimate could be strengthened and refined in the concluding section.

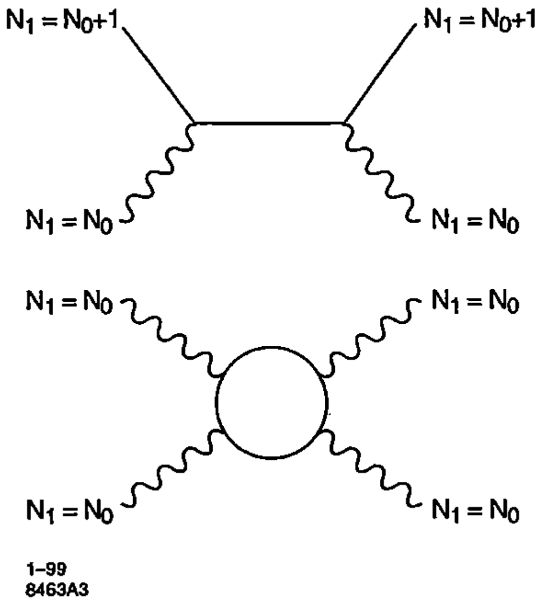


Figure 1: Comparison of bit-string labeled processes after the label length is fixed at 256 interpreted as baryon ($N_1 = N_0 + 1$) photon ($N_1 = N_0$) and photon-photon scattering. Here N_1 and N_0 symbolize, respectively, the number of ones and zeros in the label part of the string (which is of length 256). Program universe guarantees that, in the absence of further considerations, the content part of the strings will have an equal number of zeros and ones with very high probability as the string length (universe) grows.

3.4 Comparison with Observation

The currently accepted way to set the stage for the discussions of cosmology is to note that if the universe is homogeneous and isotropic on a large enough scale (for which hypothesis there is now claimed to be good evidence) and postulate Einstein gravitation (at least in the weak field limit, for which there is again good evidence) the Friedman-Robertson-Walker (FRW) equations apply. Further, if we know the boundary conditions at the time of big-bang nucleosynthesis (when the event horizon was "only" a million or so times smaller than it is now), we can integrate these equations up to the current time knowing only the two parameters Ω_M and Ω_Λ [22]. Here $\Omega_M = \rho_M/\rho_c$ is the ratio of the contemporary density of matter to the critical density in the absence of a cosmological constant ($\Omega_\Lambda = 0$). We can get this knowing only Newton's gravitational constant G_N and the current value of the Hubble constant $H_0 \equiv 100\, h_0\ km\ s^{-}1Mpc^{-}1$ since $\rho_c = 3H_0^2/8\pi G_N = 1.88\times10^{-29}h_0^2\ g\ cm^{-3}$. Similarly the scaled cosmological constant is given by $\Omega_\Lambda = \Lambda c^2/3H_0^2$, where Λ is the integration constant which must appear in solving the FRW differential equations.

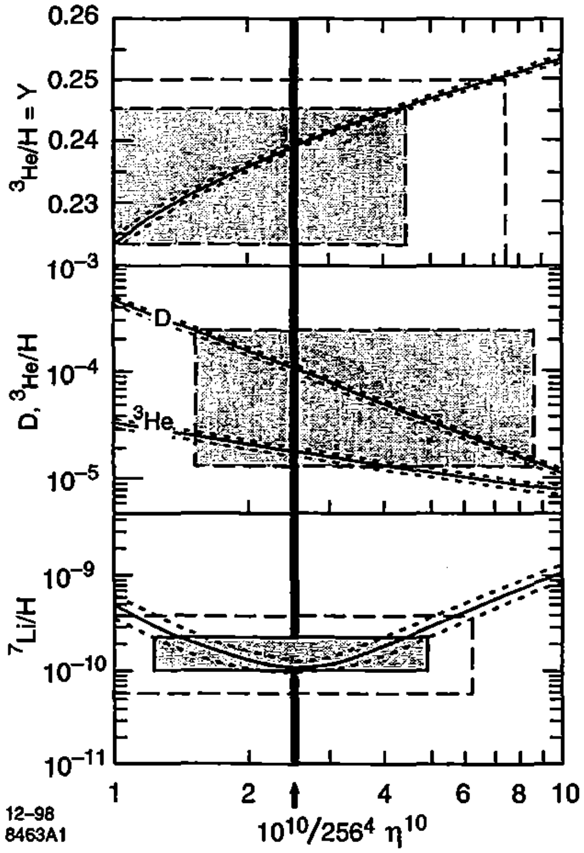


Figure 2: Comparison of the bit-string physics prediction that $\eta = 256^{-4}$ with accepted limits on the cosmic abundances as given by Olive and Schramm in [23], p. 119.

The two *program universe* results we consider here are that the ratio of dark to baryonic matter is 12.7 to 1 and that the ratio of baryons to photons at the time of nucleosynthesis, symbolized by $10^{-10}\eta_{10}$, is $1/256^4$. Our naive arguments for these numbers are given in the last section; from now on we accept them as predictions to be tested. We show in Fig. 2 that this value for η_{10} well represents a central value consistent with the cosmic abundances of the light elements. Since we know from

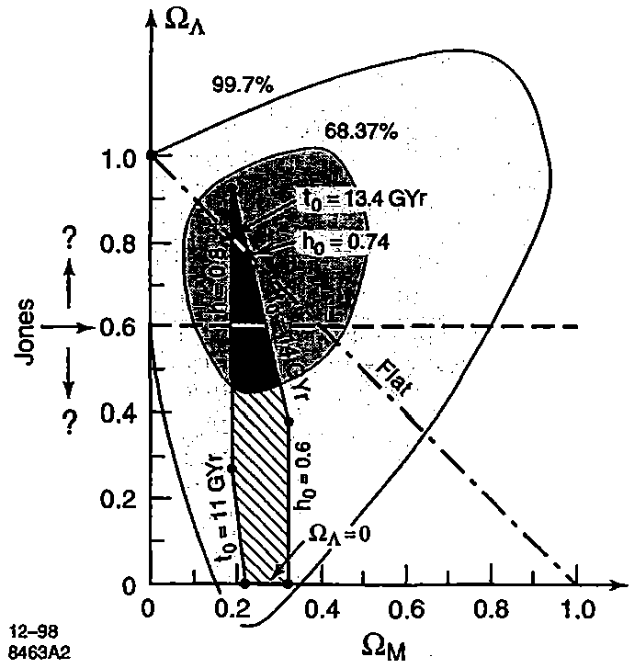


Figure 3: Limits on $(\Omega_M, \Omega_\Lambda)$ set by combining the Supernovae Type Ia data from Perlmutter, et al. with the Cosmic Ray Background Experiment (COBE) satellite data as quoted by Glanz [14] (dotted curves at the 68.37% and 99.7% confidence levels) with the predictions of bit-string physics that $\eta_{10} = 10^{10}/256^4$ (cf. Fig. 1) and $\Omega_{\text{Dark}}/\Omega_B = 12.7$. We accept the constraints on the scaled Hubble constant $h_0 = 0.7 \pm 0.1$ [5] and on the age of the universe $t_0 = 12.5 \pm 1.5$ Gyr (solid lines). We include the predicted constraint $\Omega_\Lambda > 0$). The Jones estimate of $\Omega_\Lambda = 0.6$ is indicated, but the uncertainty is not available.

the currently observed photon density (calculated from the observed 2.782 ^oK cosmic background radiation) that the normalized baryon density is given by [22]

$$\Omega_B = 3.67 \times 10^{-3} \eta_{10} h_0^{-2} \tag{2}$$

and hence from our assumption about dark matter that the total mass mass density will be 13.7 times as large, we have that

$$\Omega_M = 0.11706 h_0^{-2} \ . \tag{3}$$

Hence, for $0.8 \geq h_0 \geq 0.6$ [6], Ω_M runs from 0.18291 to 0.32517. This clearly puts no restriction on Ω_Λ.

Our second constraint comes from integrating the scaled Friedman-Robertson-Walker (FRW) equations from a time after the expansion becomes matter dominated with no pressure to the present. Here we assume that this initial time is close enough to zero on the time scale of the integration so that the lower limit of integration can be approximated by zero [24]. Then the age of the universe as a function of the current values of Ω_M and Ω_Λ is given by

$$\begin{aligned} t_0 &= 9.77813 h_0^{-1} f(\Omega_M, \Omega_\Lambda) \ Gyr \\ &= 9.77813 h_0^{-1} f(0.11706 h_0^{-2}, \Omega_\Lambda) \ Gyr \end{aligned} \tag{4}$$

where

$$f(\Omega_M, \Omega_\Lambda) = \int_0^1 dx \sqrt{\frac{x}{\Omega_M + (1 - \Omega_M - \Omega_\Lambda)x + \Omega_\Lambda x^3}} \ . \tag{5}$$

For the two limiting values of h_0, we see that

$$\begin{aligned} h_0 &= 0.8, \ \ t_0 = 12.223 f(0.18291, \Omega_\Lambda) \ \ Gyr \\ h_0 &= 0.6, \ \ t_0 = 16.297 f(0.32517, \Omega_\Lambda) \ \ Gyr \ . \end{aligned} \tag{6}$$

The results are plotted in Fig. 3.

To orient ourselves in the $(\Omega_M, \Omega_\Lambda)$ plane, we first consider a flat universe in which the curvature term in the normalized FRW equations vanishes, i.e. $1 - \Omega_M - \Omega_\Lambda = 0$. Then for $h_0 = 0.8$, performing the integration, these constraints predict $t_0 = 13.8 \ Gyr$, barely within the allowed range. The upper limit of 14 Gyr requires that $h_0 = .737, \Omega_M = .199, \Omega_\Lambda = .801$. We conclude that requiring flatness together with $\eta_{10} = 2.33$ and $M_{\text{Dark}}/M_B = 12.7$ restricts us to the short line segment from $(\Omega_M, \Omega_\Lambda) = (0.183, 0.817)$ to $(0.199, 0.801)$. At the same time, this sets a lower bound of 0.737 on h_0, and of 13.8 Gyr on t_0. It is therefore very important for our bit-string cosmology to know how flat it has to be. Otherwise we may soon be forced to modify or abandon this approach to cosmology as the observational data improve.

If we relax the flatness assumption, but take from our model (see Section 3.1) the requirement that the cosmological constant be repulsive (space generates more space as time goes by), the predicted limits on our parameters as plotted in Fig. 3 are well within the 99.7% confidence limit given by putting together the type Ia supernovae and the COBE data [14, 15]. At the 68.37% confidence limit provided by this data we see that, if we can find a way to justify our choice for η_{10} and the ratio of dark to baryonic matter, we require the cosmological constant to lie between 0.45 and 0.94. We conclude that the bit-string cosmology is within the observational bounds, and that either a calculation of the limits on flatness or of the limits on the cosmological constant would greatly improve the predictive power of our theory.

4 Jones' Cosmological Constant

Since Jones' paper [11] is still not submitted, I am at liberty here only to quote the following sentence

> From general operational arguments, Ed Jones has shown how to start from $\sim N$ Plancktons and self-generate a universe with $\sim N'$ baryons which—for appropriate choice of N—resembles our currently observed universe. In particular it must necessarily have a positive cosmological constant characterized by $\Omega_\Lambda \sim 0.6$

We note further that Jones' general arguments a) are completely compatible with *program universe* and b) do not in themselves fix the value of N. Further, the estimate given above, which was made before and independent of the calculations reported in the last section, falls squarely in the middle of the allowed region (see Fig. 3). Clearly, pursuing the combination of these two lines of reasoning could prove to be very exciting. We indicate how this might be done in the concluding section.

5 A Research Proposal for ANPA

We believe that the above calculations amply justify our contention made in the introduction that *if* the ANPA program can be shown in a convincing way to lead to the prediction of the two parameters η_{10} and $\Omega_{\text{Dark}}/\Omega_B$ to anything like the precision that $\hbar c/e^2$ and $[M_{\text{Planck}}/m_{\text{proton}}]^2$ are given in the lowest approximation by the older triumph [20] *then* we can provide a target for the observational cosmologists to shoot at. If that happens, as observations improve, we will be either vindicated or shown to have made some fatal flaw in our reasoning. This is a much more exciting game to play than trying to show physicists that we can approximately compute numbers that they already are confident they can measure to higher precision than we can provide.

The problem is that the naive arguments given above for the numbers studied here are—even within ANPA—unlikely to be convincing to any one other than a sympathetic reader. A friendly critic would at best characterize them as "heuristic" and a less friendly critic as "hand-waving". A hostile critic will dismiss them as "wishful thinking." I readily admit that we need to do better, but fear that the amount of work needed is beyond my reach. Fortunately, most of it is precisely what needs to be done in any case, if the elementary particle end of the ANPA program is not to stagnate. I now outline a possible research strategy.

I propose that we first construct a rigorous bit-string theory for renormalized QED in the truncated version of a single particle-antiparticle mass and the first combinatorial hierarchy approximation for the fine structure constant, i.e. $1/137$. Basically, I believe this only involves putting together two pieces of the puzzle which have already been completed, and which we now discuss.

Following a suggestion of Feynman's [25], Lou Kauffman and I have shown [26] that, given as the boundary condition a rational fraction velocity between two fixed end-points in 1+1 dimensional space-time, a finite and discrete version of the free particle Dirac Equation can be solved by an appropriate collection of bit-strings pairs interpreted as "random walks". Hence, once we have shown how to couple the beginning- and end-points to bit strings representing photons which satisfy the appropriate conservation laws in three dimensions, the "renormalized single particle propagator" problem for fermions will have been solved [27].

Again following an idea originally due to Feynman [28], as presented by Dyson [29] and developed further by Tanimura [30], Kauffman and I have shown [31] that the discrete physics hypothesis that first measuring position and then velocity is different from first measuring velocity and then position leads to the *relativistic* commutation relations needed to undergird, rigorously, the Feynman-Dyson-Tanimura "proof" of the free particle Maxwell Equations using in addition only Newton's second law connecting force to field. This amounts to (for a single particle trajectory) emission or absorption of "photons" at finite and discrete points in 3-space connected by straight line segments. We have noted above that these in turn can be represented by collections of random walks of a Dirac particle. What remains is to show that the "interaction" so described does indeed consistently describe the connection between bit-string photons and bit-string Dirac particles in the $\hbar c/e^2 = 137$ third level hierarchy calculation. That this way of looking at the hydrogen atom also provides the *relativistic* connection between binding energy, principle quantum number and coupling constant first given by Bohr [32] has already been proved [10]. Bits and pieces of the geometrical interpretation of the angles between bit-strings needed to lace all this together also exist [3].

I feel that a concerted effort could get to a lowest order renormalized QED in this way, providing the bit-string underpinning for the renormalized Feynman diagrams needed to discuss the equilibrium between protons and black body gamma radiation (Compton scattering and photon-photon scattering) to a part in 137 at the time of "big-bag nucleosynthesis". The black-body spectrum itself is guaranteed by the statistical character of string creation in program universe and the indistinguishability (in the usual sense of Bose-Einstein statistics) of the photon states in a bit-string representation of photons. If the bit-string version of the finite particle number relativistic scattering theory we have started to construct [18] and the quantum numbers of the standard model we have sketched [9] are not sufficient to describe the nuclear physics to the level needed at the time of big bang nucleosynthesis and later, our cosmology obviously cannot get off the ground. This is the reason why, up to now, I have given priority to putting the elementary particle physics and scattering theory on a firm foundation.

The next step, as I see, it is to make a more careful analysis (or possibly to run computer experiments) to find out reliably—rather than heuristically—what the probable distribution of label and content strings generated by program universe must be. This may actually help in getting the quantum number interpretation of the bit-string scattering theory straight. If this does *not* end up giving something close to $1/256^4$ for η, either program universe will have to be modified, or the whole bit-string cosmology abandoned.

These steps in turn are needed—but presumably at a much earlier stage in the string evolution described by program universe than we have been discussing above—in order to gain confidence in the prediction of 12.7 for the dark matter-baryon ratio. Here I foresee two ways to go. One is to revive an old idea of Wheeler's [33]: *Geons.* These are classical configurations of electromagnetic waves which are so energetic that their mass is sufficient to bind them together gravitationally as standing waves. Within classical physics, Wheeler showed that they are indeed stationary solutions of the coupled Einstein and Maxwell equations in the absence of particles. Because they are classical, they can be of any size thanks to scale invariance; the only dimensional constants which occur in the theory are G_N and c. But once one includes Planck's constant, breaking scale invariance, Wheeler showed that quantum effects start to become important already when the masses are many times the range of stellar sizes, and rapidly become dominant at smaller scales. Thus there were no observed candidates for such objects when the paper was published. But now that we know that there are enormous distributions of Dark Matter of the size of clusters of galaxies [7], we do have observational evidence that might be relevant.

The problem, as with particulate dark matter (which we discuss below), is to see how program universe might be expected to generate such structures. In the completed label scheme the string which interacts with *everything* is the unique label of length 256 which contains 256 ones (the *anti-null string*). This will represent the Newtonian static gravitational interaction. The combinatorial hierarchy construction shows that for protons this corresponds to a dimensionless coupling constant $G_N m_p^2/\hbar c \approx 2^{-127}$. But at earlier levels in the construction the analogous anti-null string occurs with probabilities $1/3, 1/10, 1/137$ as levels 1, 2 and 3 are completed. Thus the equivalent of a very strong gravitational interaction occurs during very early stages of the construction. This will bind together electrically neutral objects, some of which will continue to be electromagnetically neutral as the strong, electromagnetic, and weak interactions evolve toward their final form. These early objects might end up as something like enormous geons as the construction proceeds. The idea looks to me to be work exploring both in classical and in bit-string physics.

For the particulate dark matter, we also have a class of candidates. In our unsuccessful attempt to get our views into *Physical Review Letters* [34], we pointed out that a proton together with 2^{127} gravitating proton-antiproton pairs assembled with spin 1/2 within a radius of $\hbar/m_p c$ would form a "charged, rotating black hole" with Beckenstein number 2^{127}. It would then rapidly decay by Hawking radiation, but in our theory, since baryon number is conserved, would leave behind a proton. Therefore we claim that in our theory the proton is "gravitationally stabilized". Although we did not point it out in that note, the same argument stabilizes an electron, due to charge and lepton number conservation, and also stabilizes a (massive) electron-type neutrino, due to lepton number conservation. The neutral current interaction will, of course, provide the neutrino with a finite mass once the label-content assemblage has developed far enough. So our theory will provide neutral assemblages of photons, gravitons and (electron-type) neutrinos which bid together gravitationally. These will be our candidates for particulate dark matter.

In both cases we need to (a) work out the actual models for this neutral dark matter and (b) show that the 127/10 argument does apply to them when we have studied program universe in more detail. For the particulate types, we will also be under the obligation to calculate detection cross sections and show that extant dark matter searches would not have picked them up. Of course, if we are very lucky, we might be able to suggest new types of searches that would pick up our candidates, if they are there.

To complete the task we must, minimally, show in detail how the Jones argument (cf. Section 4) applies to program universe. This does not appear to be too difficult. Better, by examining program universe in more detail we might provide a statistical law as how the ratio between space and matter evolves. This could then form the basis for an actual calculation of the cosmological constant in the most probable of all universes. Whether this is also the best of all universes we will leave to the theologians to argue.

I hope I have said enough to indicate some of the exciting things that attention to cosmology could open up for the ANPA program in the future. I can only hope that I will be around long enough to see some of them bear fruit.

In closing I wish to thank Ed Jones for several illuminating discussions of cosmology before and after ANPA 20, and Brian Koberlein for checking out my understanding of the implications of current observations at ANPA 20 before I made my presentation. Of course I am responsible for any errors that may have crept in.

References

[1] A. F.Parker-Rhodes, "Hierarchies of Descriptive Levels in Physical Theory", Cambridge Language Research Unit, internal document I.S.U.7, Paper I, 15 January 1962. Included in K. Bowden, ed. " General Systems and the Emergence of Physical Structure from Information Theory", special issue of the *International Journal of General Systems* (in press).

[2] A. F. Parker-Rhodes, *The Theory of Indistinguishables*, *Synthese*, **150**, Reidel, Dordrecht, Holland, 1981.

[3] H. P. Noyes, "A Short Introduction to Bit-String Physics", in *Merologies: Proc. ANPA 18*, T. Etter, ed., July 1997, pp. 21-61; available from ANPA c/o Professor C. W. Kilmister, Red Tiles Cottage, High Street, Bascombe, Lewes, BN8 5DH, United Kingdom.

[4] See [3], Sect. 2, pp 28-32, for a brief discussion of *program universe* and a guide to the literature.

[5] D. E. Groom, et al., "Astrophysical Constants", in [23], p.70.

[6] C. J. Hogan, "The Hubble Constant", in [23], pp 122-124.

[7] T. Tyson, "Dark Matter Tomography", SLAC Summer Institute, 1998 (available from SLAC website).

[8] H. P. Noyes and D. O. McGoveran, *Physics Essays*, **2**, 76-100 (1989); p.96: "...This prediction is in agreement with observation, since the observed visible

mass within the event horizon is about what we estimate.... With such an open universe, Newtonian gravitation is good enough for post-fireball calculations." p. 97: "The prejudice of most cosmologists is that the universe should be closed...(I find an open universe much more satisfactory...)"

[9] H. P. Noyes, "Bit-String Physics, a Novel 'Theory of Everything' ", in *Proc. Workshop on Physics and Computation (PhysComp '94)*, D. Matzke, ed.,94, Los Amitos, CA: IEEE Computer Society Press, 1994, pp. 88-94.

[10] D. O. McGoveran and H. P. Noyes, *Physics Essays*, **4**, 115-120 (1991).

[11] E. D. Jones, "Microcosmology", private communication to HPN, June 3, 1998.

[12] M. Perryman, "The Hipparcos Astronomy Mission", *Physics Today*, June 1998, pp 38-45 and references therein.

[13] B. Schwarzschild, *Physics Today*, June 1998, pp 17-19 and references therein.

[14] J. Glanz, *Science*, **280**, 1008, 15 May 1998.

[15] J. Glanz, *Science*, **282**, 1247, 13 September 1998.

[16] M. J. Manthey, "Program Universe" in SLAC-PUB-4008, June 1986 (Part of Proc. ANPA 7), pp 101-110.

[17] Comment of J. A. Wheeler to the author when *program universe* was first discussed with him a decade and a half ago.

[18] H. P. Noyes and E. D. Jones, "Solution of a Relativistic Three Body Problem", submitted to *Few Body Systems: Tjon Festschrift Issue* and SLAC-PUB-7609(rev) 1998.

[19] D. O. McGoveran, private communication to HPN November 19, 1998. McGoveran has tried to get this point across to this author, and to others committed to the ANPA research program for several years.

[20] T. Bastin, "On the Scale Constants of Physics", *Studia Philosophica Gandensia*, 4, 77 (1966).

[21] T. Bastin, H. P. Noyes, J. Amson and C. W. Kilmister, *Int'l. J. Theor. Phys.*, **18**, 445-488 (1979).

[22] K. A. Olive,"Big-bang Cosmology", in [23], pp 117-118.

[23] Particle Data Group, "Review of Particle Properties", *Euro. Phys. J.* **C 3**, 1-794 (1998).

[24] J. R. Primack, "Dark Matter and Structure Formation", in *Formation of Structure in the Universe*, Proc. of the Jerusalem Winter School 1996, A. Dekel and J. P. Ostriker, eds. Cambridge University Press (in Press) and astro-ph/9797285v2 25 Jul 1997,

[25] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, 1965, Problem 2-6, pp 34-36.

[26] L. H. Kauffman and H. P. Noyes, *Physics Letters*, **A 218**, 139-146 (1996).

[27] H. P. Noyes, *Physics Essays*, **8**, 434 (1995).

[28] R. P. Feynman, conversation with Dyson c. 1948, reported in [29].

[29] F. J. Dyson, *Amer. J. Phys.*, **58**, 209 (1990).

[30] S. Tanimura, *Annals of Physics*, **220**, 229-247 (1992).

[31] L. H. Kauffman and H. P. Noyes, *Proc. Roy. Soc. Lond.*, bf A, 81-95 (1996).

[32] N. Bohr, *Phil Mag.*, 332-335 (February, 1915).

[33] J. A. Wheeler, "Geons", *Phys. Rev*, **97**, 511-36 (1955).

[34] H. P. Noyes, "Comment on 'Statistical Mechanical Origin of the Entropy of a Rotating, Charged Black Hole' ", SLAC-PUB-5693 (November, 1991), unpublished.

A Noncommutative Approach to 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

1 Introduction

This paper is an expanded version of [18] 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 noncommutative algebra and noncommutative vector analysis. In this way a formalism similar to quantum mechanics occurs first, but not necessarily with the usual interpretations. By following this line carefully we can show how the outlines of the well-known commutative forms of physical theory arise first in noncommutative form. The exact relation of commutative and noncommutative theories raises a host of problems.

In the first section of this paper we discuss the properties of the noncommutative discrete calculus that underlies our work. The next section examines the consequences for a particle whose position and 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. In the next section we discuss how our discrete 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 our commutators, then the dynamical variables will obey Hamilton's equations. Thus we can take

Hamilton's equations as a classicization of our theory. The next section 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 represents one tick of the clock. Then follows a section 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 recent model for quantum gravity due to Louis Crane.

Acknowledgement. It gives the author pleasure to thank Pierre Noyes, Basil Hiley, Keith Bowden and Arleta Giffor for helpful conversations, the National Science foundation for support of this research under NSF Grant DMS-9802859 and the National Security Agency for support of the research under grant MDA904-97-1-0015.

2 Discrete Ordered Calculus

In this section we recall the construction of an ordered version of the calculus of finite differences *DOC* [14], [18]. 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

$$X, X', X'', \ldots.$$

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.

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.

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

Definition. Let

$$dX = X' - X$$

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

$$X, X', X'', \ldots$$

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 noncommutative algebra. We shall assume that the values of the time series are in a possibly noncommutative 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 noncommutative 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 noncommutative 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 [14] Pierre Noyes and the author use this foundation to build a derivation of a noncommutative version of electromagnetism. Another version of this derivation can be found in [17]. 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 noncommutative geometry [5].

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 [14] we show how the formalism of electromagnetism arises when g^{ij} is δ^{ij}, the

Kronecker delta. In [16] we will show how the general case corresponds to a "particle" moving in a noncommutative 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 [21] in a discrete context.

It should be emphasized that all physics that we derive in this way is formulated in a context of noncommutative operators and variables. We do not derive electromagnetism, but rather a noncommutative analog. It is not yet clear just what these noncommutative 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 [10].

To return to basics, consider the commutator equation

$$[X, DX] = Jk$$

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. It is not so easy, however to obtain similar solutions in higher dimensions under the restrictions we impose in the the next section. As a result, most of the rest of this paper will be rather abstract, a mathematical schema in search of realization.

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

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

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

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

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

$$\nabla_{ij} F = [X_i, [X_j, F]].$$

We apply the operator ∇_{ij} to the second DOC derivative of X_k.

Lemma 4. $\Gamma_{ijk} = (1/2)\nabla_{ij} D^2 X_k$

Proof.

$$\nabla_{ij} D^2 X_k = [X_i, [X_j, D^2 X_k]]$$

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

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

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

$$= [X_i, \partial_s g_{jk} \dot{X}^s] + [\dot{X}_k, g_{ij}] + [\dot{X}_j, -g_{ik}]$$

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

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

It is remarkable that the form of the Levi-Civita connection comes up directly from this non-commutative calculus without any apriori geometric interpretation.

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 noncommutative analog of a gauge field. G_i is a noncommutative analog of a scalar field. The details of these calculations will be found in [16].

This description of the equations for a noncommutative 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 noncommutative equations of the discrete ordered calculus.

4 Poisson Brackets and Commutator Brackets

Dirac [8] 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 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 noncommutative world, we begin these questions anew.

In making this backwards journey to classical physics we see that it is necessary to make at least one further restriction on the commutation relations. For in Poisson brackets it is the case that

$$\{q_i, q_j\} = 0$$
$$\{p_i, p_j\} = 0$$
$$\{q_i, p_j\} = \delta_{ij}.$$

In our formalism, we would identify X_i as the correspondent with q_i and $dotX^j$ as the correspondent of p_j. Thus to match the Poisson algebra we would need to demand that

$$F^{ij} = [\dot{X}^i, \dot{X}^j] = 0.$$

Under these circumstances we would have

$$D^2 X_i = G_i + \Gamma_{ijk} \dot{X}^j \dot{X}^k,$$

the usual form of the description of geodesic motion with respect to the Levi-Connection. (See the previous section for the discussion of these equations.) The non-commutative calculus structure that lies below this has a non-trival gauge field that arises from the extra commutation relation. This raises further questions about the nature of the generalization that we have made. Originally Hermann Weyl [22] 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.

5 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 [19] should be able to accommodate time. A suggestion directly related to this remark occurs in [7], and we shall take this up at the end of the next section of this paper.

6 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 [13], [1],[23]. 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 [3] 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.

6.1 Remarks on Quantum Mechanics

We should remark on the basic formalism for amplitudes in quantum mechanics. The Dirac notation $\langle A|B\rangle$ [8] 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

$$A - - - * - - - C^1 - - - * - - - C^2 - - - * - - - ... - - - * - - - C^m - - - B$$

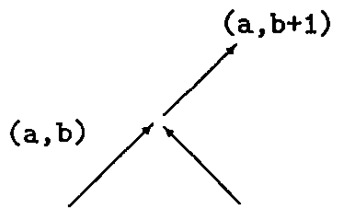
With A and B fixed the amplitude for the net is

$$< A|B >= \Sigma_{1\le i_1 \le ... \le i_m \le n} < A|C^1_{i_1} >< C^2_{i_2}|C^3_{i_3} > ... < C^m_{i_m}|B >$$

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 [12] 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 [8] in one space variable and one time variable. This example is related to an observation of Richard Feynman [12]. In [15] 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 [15]. I am indebted to Charles Bloom [2] 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 [20] by Schweber. It is still an open problem to generalize this exercise of Feynman to four dimensional discrete spacetime.

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 [9]. Concepts of time and category are discussed by Louis Crane [6], [7] 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.

6.2 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 [4] 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 [4]. 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 the Crane model. Much more work needs to be done in this domain.

References

[1] M.F. Atiyah [1990], *The Geometry and Physics of Knots,* Cambridge University Press.

[2] Bloom, Charles [1998], (private communication)

[3] R. Penrose [1971], Applications of negative dimensional tensors, In *Combinatorial Mathematics and Its Applications*, edited by D. J. A. Welsh, Academic Press.

[4] Crane, Louis , Kauffman, Louis H., Yetter, David N. [1997], State sum invariants of 4-manifolds, *Journal of Knot Theory and Its Ramifications*

[5] Connes,Alain [1990], *Noncommutative Geometry* Academic Press.

[6] Crane,Louis [1996], Clock and category: Is quantum gravity algebraic?, *J. Math. Phys.* **36** (11), November (1996), pp. 6180-6193.

[7] Crane, Louis [1997], A proposal for the quantum theory of gravity, (preprint).

[8] Dirac, P.A.M. [1968], *Principles of Quantum Mechanics,* Oxford University Press.

[9] T. Etter, Process, System, Causality and Quantum Mechanics', *International Journal of General Systems*, issue edited by K. Bowden on **General systems and the Emergence of Physical Structure from Information Theory**, (in press).

[10] E. Fredkin [1990], Digital Mechanics, *Physica D* **45**, pp. 254-270.

[11] Ashtekar,Abhay, Rovelli, Carlo and Smolin,Lee [1992], "Weaving a Classical Geometry with Quantum Threads", *Phys. Rev. Lett.*, vol. 69, p. 237.

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

[13] Kauffman,Louis H.[1991,1994], *Knots and Physics,* World Scientific Pub.

[14] 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.

[15] Kauffman,Louis H. and Noyes,H. Pierre [1996], Discrete Physics and the Dirac Equation, *Physics Letters A*, 218 ,pp. 139-146.

[16] Kauffman,Louis H. and Noyes,H.Pierre (In preparation)

[17] Kauffman, Louis H.[1996], Quantum electrodynamic birdtracks, *Twistor Newsletter Number 41*

[18] Kauffman, Louis H. [1998], Noncommutativity and discrete physics, *Physica D* 120 (1998), 125-138.

[19] 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.

[20] Schweber, Silvan S. [1986], Feynman and the visualization of space-time processes, *Rev. Mod. Phys.* Vol. 58, No. 2, April 1986, 449 - 508.

[21] 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.

[22] Weyl, Hermann [1922], *Space–Time–Matter*, Methuen, London (1922).

[23] Witten,Edward [1989], Quantum field Theory and the Jones Polynomial, *Commun. Math. Phys.*,vol. 121,pp. 351-399.

Light and Dowsing

by D.H.Wheeler

Introduction

The best evidence that dowsing is a genuine, natural phenomenon relates to water dowsing. Here dowsers are employed to find water on a 'No water – No pay' basis. Even so, they are in constant demand for their professional services, particularly from farmers. A study by Professor H-D.Betz (a physicist in the University of Munich) reviewed the search for water in Sri Lanka during the 1980's. From 691 test drillings, the dowsers achieved an overall success rate of 96% whereas conventional techniques only gave a 30-50% success rate.

Dowsing first became a serious interest for me after I read a newspaper article explaining how Professor Reddish (Emeritus Professor of Astronomy, Edinburgh University) walked over a linear object (eg. a pole or wire) with two dowsing rods in his hands (see Fig 1) and they rotated to cross over in front of his chest. He subsequently showed that walking under a suspended wire gave the same reaction. If a plastic tube was put down on the ground under the length of the wire, the rod crossing did not take place under the wire, but at intervals on either side of the plastic tube. For Professor Reddish[2] this represented an example of a classical wave interference pattern, but for me it showed that dowsing was real physics in action.

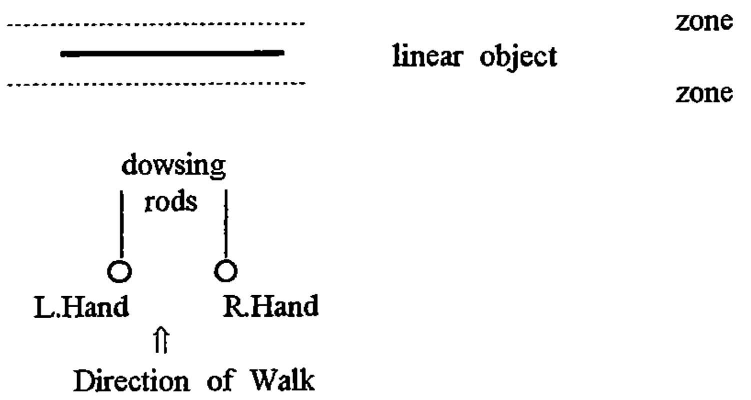


Fig 1

Trying to measure the interference patterns from multiple linear objects (wooden poles) in a field, I found that walking between two trees caused the dowsing rods to rotate. So, walking over a linear object causes a rod rotation and so does walking between two trees. Associating the trees with circular bases, I placed dinner plates on the floor of my living room and obtained the same rod rotations. Diagrammatically, the rods turned from ↑↑ to →←.

It then transpired that any object with a circular base (eg. a wine bottle or saucepan) would initiate the crossing of the dowsing rods. From this point onwards, I conducted a variety of experiments with linear and circular objects in different configurations to try and establish some basic patterns or reactions from the dowsing rods that would help lead to a better understanding and could also be duplicated by others.

Linked to our Consciousness

Nobody has yet been able to come up with an explanation of the dowsing phenomenon, but the dowsing fraternity are very firmly convinced that the mind is involved. Besides the para-normal ability called Map-Dowsing, which remotely locates missing objects like keys and people, I have come to accept that the mind is able to carry out some sort of tuning function that enables it to concentrate on the job in hand and exclude all the other extraneous background signals (eg. from other linear and circular objects around) that would otherwise be expected to cause some interference. The mind seems to be able to pick up these tuned signals and trigger a response that enables the dowsing rods to rotate.

Experiments

Practically all my experimental evidence has been obtained in the home using linear objects (eg. wooden poles, plastic tubes, wire, golf clubs) and circular ones (eg. dinner plates, plastic bowls, wine bottles). It is estimated that at least 80% of people can dowse, so readers have the opportunity to confirm the results for themselves in the comfort of their own living rooms.

General Observations on what can be measured.

1. *Materials*. The dowsing phenomenon occurs both with ferro-magnetic (eg. steel, iron) and non ferro-magnetic materials (eg. wood, bone, plastic).

2. *The shapes or boundaries* of objects are important in determining where the dowsing rods rotate and the locations seem to follow the rules of Euclidean geometry. That is, rotations occur when the dowser arrives along a track at right-angles to:

(a) the boundary of a linear or circular object.

(b) the imaginary line connecting the centres of circles

(c) the imaginary line from a linear object or perpendicular from it, which intersects with the centre of a circle.

(d) the extended imaginary line that connects other linear objects

(e) the 'external' imaginary line which bisects the apex of linear objects forming an isosceles triangle

I term these rotation points the Euclidean Locations for Rotation.

3. *All objects can affect all other objects. A pressure effect?*

Linear and circular objects are found to interact with one another in producing dowsing rod rotations. As everything can be considered to be made up from lines and curves, it is but a small step to consider that all objects can potentially interact with any other object. As a speculation, this interaction can be regarded as being similar to Einstein's second component of gravity, the pressure that matter exerts upon its surroundings (λ lambda, the cosmological constant), which can be positive or negative..

3. *The dowsing zone* where rod rotations commence and end is quite narrow; inches rather than feet wide.

4. The *strength* of the rod rotation (as felt in the hands) is generally very similar for all experiments, but it has been known to vary. It can be particularly strong when dowsing with Y-shaped twigs rather than L-shaped rods. Adding one object to another does not increase the strength of the rod rotation, but the effect may possibly alter the shape of the combined object and its Euclidean Locations for Rotation.

5. *Electromagnetic Association.* Walking through a light beam will initiate a dowsing rod rotation. Other researchers findings[3,4,5] show that electromagnetic fields such as radio waves, ionising radiation, electric fields and electrostatic shocks either affect, or are affected by, dowsing zones. This evidence suggests that there is a dowsing field (D-Field) present, which can co-exist with electromagnetic radiation, although it is not electromagnetic itself.

6. D-Fields (Dowsing Fields)

There are a number of dowsing experiments that provide further indications for the existence of D-Fields.

(i) Interference Patterns

By varying the height of his suspended wire above the plastic tubing on the ground, Reddish[1] found that the intervals between the dowsing zones on the ground increased as the wire height decreased (and vice-versa).

These variations in the interference patterns suggest field effects.

(ii) Rod Angles

The angles rotated by the rods vary, but are commonly found to be 0° (↑↑), 45°(↗↖), or 90°(→←). Other intermediate values and negative values are found (eg. ↓↓, ↙↘, ←→).

(iii) Universal Field

The rods rotate back to their forward direction after leaving the dowsing zone.

Objects are found to leave imprints or memories of their position long after they have been removed[2,4,6]. This imprinting can last for minutes, days or years.

Both of these effects suggest the presence of a Universal Dowsing Field (UD-Field).

(iv) Motion / Inductance

Raising the dowsing rods quickly into the horizontal search position, and/or walking briskly into the dowsing zone seems to enhance the sensitivity of the dowsing rod reaction. Such a result brings to mind the concept of 'cutting lines of force'. The same type of result was obtained by Tromp[5] using a changing magnetic field, which could be detected and a static one, which could not.

If the D-Field can exhibit effects normally associated with electromagnetism, it is intriguing to speculate on where are the induction effects both on earth and the cosmos.

7. Body Receptors and Vision.

The human body has receptors for the external D-Fields. In 1978, Zarbor Harvalik,USA carried out an experiment with a low power HF generator, to which dowsers responded with the normal dowsing reaction ie. the rods crossed, when they walked through the HF beam. By inserting screening bands (copper or aluminium) between the dowser and the generator, he found that dowsing reactions were prevented at the level of the adrenal glands and the pineal gland.

For me, the eyes and vision are essential as I cannot dowse in the dark or with my eyes closed. Experiments provided intriguing results where, for example

(a) Using two rods to dowse over a linear wooden pole, if the left eye was kept open only the left rod rotated and the right rod stayed pointing forward. Similarly if only the right eye was kept open.

(b) In a dark room where no dowsing reaction was observed, switching on a spot light/torch light beam and walking through it, induced the normal dowsing rod crossing reaction.

(c) An independent observer looking down the line of the wooden pole, prevented my getting any dowsing reaction. This suggests that another person's vision or thought waves are able to inhibit my sensors or my mind's ability to react to the external field effect.

(d) No dowsing reaction was obtainable unless the rods were kept in view.

Whilst it would be satisfying to report that no one else can dowse in the dark, this is not the case and many people can. However, light and vision do seem to be important for me (and others) to get the dowsing reaction.

Why do the Rods Rotate?

Where does the rod rotational energy come from?

Opinion is divided upon whether it is a field effect or whether the rods rotate under the action of gravity and involuntary muscular action. I consider the field effect has some part to play, because I cannot see or feel the muscular activity and some dowsers using the Y-rod have experienced the bark stripping away from their wooden rod as a result of their efforts to resist the twisting action in their hands. But, if there is a field effect, is the energy supplied externally or from the human body or mind? Perhaps it's all three.

Measuring Brain Waves

Experiments in the USA with an expert dowser connected up to a Brainwave Analyser have shown that:-

(i) the dowser has an expanded brain wave state that is different from transcendental meditation or dreaming states.

(ii) there is simultaneous high power levels of beta, alpha, theta and delta waves, with all the neurons on both sides of the brain firing in synchronicity ie. highly coherent brain waves.

Such results cause the researchers to claim that the dowser was in a truly altered state of consciousness.

Opportunities for Measurement

There are many physical measurements that can be made related to dowsing. But why should anyone bother to carry them out? The answer I like to believe is that by doing so, the results will provide evidence for physical fields that science is reluctant to accept, yield more information about the way the mind processes information and give clues towards a new theory of quantum mechanics. Such new information will hopefully contribute towards a better understanding of how the conscious and sub-conscious minds interact together, and the impact they have on nature and the environment.

The dowsing phenomena is one of the easiest para-normal effects to measure and investigate, as it only requires low technology and simple facilities to carry out the research work. Unfortunately it seems to have been ignored because of scepticism and disbelief, in spite of the highly convincing evidence from professional dowsers[1] and other impressive published accounts[4,6].

Reproducibility *Experiments on the Web site.*

Science demands that experiments should be reproducible before the results can be considered to have any value. To this end, I have arranged for five simple dowsing experiments to be put on the Scientific & Medical Web site http://www.cis.plym.ac.uk/SciMedNet/library/articles/9712071705.

I invite readers to try these for themselves and send their results back by e-mail or letter under the heading 'Dowsing Experiments'. These experiments only require the dowser to comment on the rod rotations at specified locations using simple poles, golf clubs, tubes, plates, or discs on the floor. Instructions are given for beginners. This is a serious request and the results will be put into a data base for future reference.

References

1. Applegate, George *The Complete Guide to Dowsing* Element Publ 1997

2..Reddish, V.C., *The D-force - A remarkable phenomenon* Jane Street Print Co, Edinburgh (1993)

3 Maby J.C. and T.Bedford Franklin *Physics of the Divining Rod* Bell 1939

4. G.P.Hansen, *Dowsing: a review of experimental research* Journal of the Society of Psychical Research 1982, Vol 51 No.792, pp343-367.

5. Tromp S.W *Psychical Physics* Elsevier Publishing 1949

6. Bird, Christopher *The Divining Hand* Whitford Press 1993

Crossing Symmetry and the Equivalence Principle in Einsteinian Gravity

Brian D Koberlein
Department of Physics, SUNY Empire State College,
1475 Winton Rd. N., Rochester, NY 14609

Crossing symmetry, when applied to the gravitational problem, implies a repulsion between matter and antimatter, in direct violation of the principle of equivalence. However, this is not as problematic to Einstein's theory as one might think. This work begins with a discussion of the relation between crossing symmetry and the more commonly applied CPT invariance. An overview of the equivalence principle as stated by Einstein is presented, and the ways in which this principle has been interpreted is discussed. Finally, positing crossing symmetry as the fundamental invariance principle, a modified form of Einsteinian gravity is presented and discussed.

ANPA 20 September 1998

I. INTRODUCTION

The Noyes prediction in discrete physics of a gravitational repulsion between matter and antimatter has been highly debated. The controversy arises because such a prediction would seem to shatter two strongly held keystones of modern physics: the equivalence principle of general relativity, and the CPT Theorem of particle physics. Despite these debates, however, the author has not seen a detailed examination of the prediction and its effects on current physical models.

The purpose of this paper is to present such an examination. In Section II, the natures of CPT invariance and crossing symmetry are discussed. It is shown that crossing symmetry does not violate the CPT Theorem but rather extends its applicability. Section III examines the principle of equivalence. It is shown that there is both a weak and strong equivalence principle. It is then argued that while the Noyes prediction violates the weak principle, it does not violate the strong. In Section IV, these concepts are applied to Newtonian Gravity. It is demonstrated that a gravitational model satisfying the Noyes prediction can be obtained if one allows for an interaction-dependent gravitational field. Finally, Section V examines the effects on Einsteinian gravity. It is shown that a discrete gravitational theory can be derived which agrees with both general relativity and the Noyes prediction.

II. CPT AND CROSSING SYMMETRY

Central to modern physics is the concept of symmetry. For example, the momentum of a closed system is conserved because the spatial coordinates are symmetrical in that they are homogeneous and isotropic. Symmetry has been used heavily in particle theory, where the concept allows one to describe particle interactions in terms of mathematical groups. In particle theory, the central concept of symmetry is the CPT Theorem, which states that the laws of physics are invariant under the combined operations of Charge conjugation, Parity inversion, and Time reversal in any order. [1]

Time reversal is simply a $t \to -t$ substitution, which results in a reversal of the direction of time. Watch a film of two interacting particles, then watch the film in reverse. If the process is invariant under T, then it is impossible to determine for which direction the film is correct. Both directions are physically possible.[1] The electromagnetic, strong, and (classical) gravitational interactions are

[1]The concept of time reversal in this context is meant to consider only simple particle interactions where the entropy of the system is constant. For large, complex systems, entropy gives a clear "arrow of time" via the second law of thermodynamics.

each T invariant. There is indirect evidence that the weak force is not T invariant, which will be elaborated upon in a moment.

Parity inversion is the spatial equivalent of Time reversal. That is, a $(x, y, z) \rightarrow (-x, -y, -z)$ substitution. It is often viewed as change of handedness, since such a substitution changes a right-handed coordinate system to a left-handed one, and vice versa. Consider a particle system placed next to a mirror. For P invariant processes, it is impossible to distinguish the system from the image as both are physically possible. As with T, all but the weak interaction are P invariant. Perhaps the most famous demonstration of P violation in weak interactions is the cobalt-60 experiment, which showed beta decay to be preferentially left handed. [2]

Finally, Charge conjugation is simply a substitution of all charges with their anti-charge counterpart.[2] It is often viewed as the mutual substitution of particles and antiparticles. A good example of C invariance is that of atomic structure. Physically, it does not matter if the electron is taken to be negative or positive, so long as the protons are of the opposite sign. Both the electromagnetic and strong interactions are C invariant. The weak interaction demonstrates C violation, again through the cobalt-60 experiment, which demonstrates that the neutrino is left handed whereas the anti-neutrino is right handed. Due to the weakness of the gravitational interaction and the lack of antimatter[3] in quantity, it is not yet known whether gravity is C invariant. There are, however, theoretical arguments in its favor.

Although C, P and T (as well as CP and PT) are each violated for some physical interactions, there is no known process which violates CPT. In addition to the lack of experimental CPT violation, there is strong theoretical support for CPT. In order for a quantum field to possess the necessary qualities of being causal, Lorentz invariant, and Hermitian, the field must be CPT invariant. Since the weak interaction exhibits CP violation through neutrino asymmetry, CPT dictates it must also violate T.

For gravity, the CPT Theorem requires that, like matter, antimatter particles should be mutually attractive. It leaves open the question of whether a particle-antiparticle pair would attract or repel. Either possibility is acceptable to CPT. However, the general opinion among theorists is that matter and antimatter should mutually attract. In quantum field theory, this view stems from the spin-2 nature of the graviton. In order for two like masses to attract one another, the field quantum for gravity must be a spin-2 boson. Such spin-2 bosons cause attraction not only for like masses but also for unlike masses. However, the details of this argument [3] rely on the *assumption* that the gravitational field is gauge invariant, which is arguably not well defined. The more common source of this view stems from Einstein's equivalence principle, which implies that matter and antimatter move in the same way for a given gravitational field. As we shall see, though, the principle of equivalence is not as strong as one is led to believe.

Since the matter-antimatter question arises from the limitations of the CPT Theorem, one would hope the question could be answered from a symmetry which is more general than CPT. Although there is no clear path in conventional physics, in discrete physics the path is clearly indicated by crossing symmetry.

Crossing symmetry, or Amson invariance [4], is a mathematical property of binary discrimination. That is, the discrimination operator $\oplus$ is usually defined by

$$0 \oplus 0 = 0;\ 0 \oplus 1 = 1;\ 1 \oplus 0 = 1;\ 1 \oplus 1 = 0, \tag{1}$$

but discrimination could equally be defined by the dual operator $\oplus'$, such that

$$0 \oplus' 0 = 1;\ 0 \oplus' 1 = 0;\ 1 \oplus' 0 = 0;\ 1 \oplus' 1 = 1. \tag{2}$$

---

[2]Charge here is taken to mean not only electromagnetic charge but also the color charges of the strong interaction. Charge conjugation can be extended to gravity through the idea of gravitational "charge." However, this concept is not well understood in conventional models, which is a central point of this paper.

[3]Some have argued that via the equivalence principle, conjugation of gravitational charge should also lead to a conjugation of inertial mass. Thus, for gravity C should apply not to antimatter but rather a "negative" matter. The properties of negative matter are even more unusual that those discussed here.

Essentially, Amson invariance amounts to a reversal $(0,1) \rightarrow (1,0)$ of the bit notation. Thus for a particular bit-string,

$$A = 1\,0\,0\,1\,1\,1\,0\,1\,0\,0\,\cdots, \tag{3}$$

$$A' = 0\,1\,1\,0\,0\,0\,1\,0\,1\,1\,\cdots. \tag{4}$$

Since each represents the same mathematical pattern, one is equally justified in using either form in the mathematical description of particle interactions in the discrete physics model.

Noyes [5] has demonstrated that if one defines discrimination with the anti-null string $\overline{I}$ as distinguishing particles from anti-particles, then particles and anti-particles are simply Amson duals. Furthermore, if a mathematically symmetric definition of the space-time coordinates is used, then the application of crossing symmetry is physically equivalent to the application of CPT. The mathematical requirement that bit-strings be Amson invariant therefore dictates that physical interactions in the discrete model be CPT invariant.

The advantage of the bit-string approach over traditional arguments is that bit-strings allow one to study particle-antiparticle interactions (including gravitational interactions) in a direct way. As Noyes and Starson have shown [6], the gravitational interaction in the Newtonian limit differs from the electromagnetic interaction only in that like masses attract rather than repel. Therefore unlike masses must repel rather than attract. In discrete physics then, Amson invariance implies that physical interactions are CPT invariant, *and* matter and antimatter should gravitationally repel.

III. THE EQUIVALENCE PRINCIPLE

When Einstein [7] first proposed the principle of equivalence as the foundation to general relativity, his basic argument was that, without some external point of reference, a free-floating observer far from gravitational sources and a free-falling observer in the gravitational field of a massive body each have the same experience. Likewise an observer standing on the surface of a massive body and an observer which uniformly accelerates at a rate equal to the body's surface gravity have identical experiences.[4] Thus, the free-float and free-fall frames can be considered equivalent. In the same manner, the uniform acceleration frame and the surface frame are equivalent. This is known as the weak equivalence principle, which can be concisely stated as: [8]

> All effects of a uniform gravitational field are identical to the effects of a uniform acceleration of the coordinate system.

In order to formulate general relativity in terms of general covariance, Einstein later strengthened this argument to yield what is known as the strong equivalence principle: [9]

> The ratio between the inertial mass of a particle and its gravitational mass is a universal constant.

It is this principle which was experimentally validated by the classic Eötvös experiment [10], which determined that objects fall at the same rate, regardless of their material consistency.

The equivalence principle stands at the core of general relativity. Its violation would forbid a metric description of gravitation, and there is a great deal of evidence, both experimental and theoretical, that gravity is a metric theory. For this reason, physicists are quick to dismiss any theory which proposes a violation of the equivalence principle. But one must be cautious in applying the principle

---

[4] Actually, the frames can be distinguished. A large enough observer would feel tension forces in the free-fall and surface frames due to the nonuniformity of the gravitational field and would not feel such tensions in the free-float and uniform acceleration frames. Here, it is assumed that the observer is small enough that tension forces are essentially undetectable.

of equivalence in such a casual fashion. What the principle actually says, and how it is generally interpreted, are often contradictory.

The popular view of the principle of equivalence is that it strictly requires all objects to fall in exactly the same way. Formally, that all objects must traverse a geodesic path in space-time. This is actually a consequence of the weak equivalence principle, which does not hold in all cases. Strictly speaking, the weak equivalence principle applies only to point masses which possess no charge and no spin.[5] It is only the strong equivalence principle which applies to point particles which have spin and/or charge.

At present, there is no experimental evidence which proves such particles must move along a geodesic. Rather, there is strong theoretical evidence to the contrary. DeWitt and Brehme [11] have shown for a point charge in a general gravitational field (one which is not globally uniform) the electromagnetic field produced by the charge is scattered by the "bumps" of the gravitational curvature. The scattered radiation then back reacts with the charge, causing it to deviate from a geodesic path. Papapetrou [12] has demonstrated that a coupling occurs between the intrinsic spin of a point particle and the local curvature of the metric, which also yields a deviation from the geodesic. Thus it is clear that violations of the weak equivalence principle occur not only in the discrete approach to gravity but also in general relativity itself.

The difficulties which arise from an equivalence principle violation are only serious if the strong principle is violated. The inability to apply the weak principle makes matters much more complicated, but such a transgression does not alter the keystone of gravitational theory. It would seem then, that to examine the validity of crossing symmetry when applied to gravity, one needs to answer a central question: Is there a way to formulate a gravitational model in agreement with Amson invariance which does not violate the strong equivalence principle?

IV. MODIFIED NEWTOWNIAN GRAVITY

In order to answer this question, we will first examine the Newtonian approximation of the gravitational field. To simplify matters, we will consider only point masses which are both uncharged and spin free. The question in this case is whether a consistent form of Newtonian gravity can be derived which is both Amson invariant and strongly equivalent.

When dealing with gravitational interactions, there are actually three different types of mass which are introduced: [13]

1. Inertial mass, m_i, which determines the inertia of an object;
2. Passive gravitational mass, m_p, which interacts with the local external gravitational field;
3. Active gravitational mass, m_a, which creates the external gravitational field in which particles interact.

Newton made the assumption that $m_i = m_p = m_a$. However, the strong equivalence principle only relates m_i and m_p. Specifically,

$$m_p = \alpha m_i \tag{5}$$

where α is a universal constant. It is generally assumed that Newton's assertion is correct, but nothing in general relativity requires it, and there is no experimental evidence to validate it. One can, however, relate m_p and m_a by imposing conservation of momentum on the system.

Consider two gravitationally interacting particles, (1) and (2), located at the position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively. The equations of motion for these particles are

[5]Since all elementary particles have both spin and charge, it is arguable that the weak equivalence principle never truly holds.

$$\overset{(1)}{m}_i\,\ddot{\mathbf{r}}_1 = G\frac{(\mathbf{r}_2-\mathbf{r}_1)\,\overset{(1)}{m}_p\overset{(2)}{m}_a}{|\mathbf{r}_2-\mathbf{r}_1|^3} \tag{6}$$

$$\overset{(2)}{m}_i\,\ddot{\mathbf{r}}_2 = G\frac{(\mathbf{r}_1-\mathbf{r}_2)\,\overset{(2)}{m}_p\overset{(1)}{m}_a}{|\mathbf{r}_1-\mathbf{r}_2|^3} \tag{7}$$

where $\ddot{\mathbf{r}} = d^2\mathbf{r}/dt^2$ and G is the gravitational constant. Addition of these equations yields

$$\frac{d}{dt}(\overset{(1)}{m}_i\,\ddot{\mathbf{r}}_1 + \overset{(2)}{m}_i\,\ddot{\mathbf{r}}_2) = G\frac{(\mathbf{r}_2-\mathbf{r}_1)\,(\overset{(1)}{m}_p\overset{(2)}{m}_a - \overset{(2)}{m}_p\overset{(1)}{m}_a)}{|\mathbf{r}_2-\mathbf{r}_1|^3}. \tag{8}$$

Since momentum is conserved for the system, the left side vanishes. It must therefore be that

$$\overset{(1)}{m}_a = \beta\,\overset{(1)}{m}_p;\;\; \overset{(2)}{m}_a = \beta\,\overset{(2)}{m}_p \tag{9}$$

where β is a constant. This requirement can also be shown to satisfy conservation of energy. Thus, the strong equivalence principle and conservation of energy-momentum dictates

$$m_p = \alpha m_i;\;\; m_a = \beta m_p. \tag{10}$$

These relations can be simplified by noting that the magnitudes of α and β can be absorbed into the gravitational constant G. It should be noted, however, that α and β can each be either positive or negative. The only condition is that α and G must be of the same sign in order to agree with experimental observations of matter interactions. Taking G to be positive, the condition of strong equivalence becomes

$$m_i = m_p = \pm m_a. \tag{11}$$

Substituting this condition into equations (6) and (7) and simplifying, the equations of motion for the particle interaction become

$$\ddot{\mathbf{r}}_1 = \pm G\frac{(\mathbf{r}_2-\mathbf{r}_1)\,\overset{(2)}{m}_i}{|\mathbf{r}_2-\mathbf{r}_1|^3} \tag{12}$$

$$\ddot{\mathbf{r}}_2 = \pm G\frac{(\mathbf{r}_1-\mathbf{r}_2)\,\overset{(1)}{m}_i}{|\mathbf{r}_1-\mathbf{r}_2|^3}. \tag{13}$$

If $m_a = +m_i$, then the particles are mutually attracted. If $m_a = -m_i$, then they are mutually repelled. Applying crossing symmetry to the interaction, we have the requirement

$$m_a = +m_i \text{ for like masses} \tag{14}$$

$$m_a = -m_i \text{ for unlike masses.} \tag{15}$$

It is clear that this represents a gravitational model which is both Amson invariant and strongly equivalent.

There is, however, a subtlety to the m_a constraint which is not evident in the simple two-particle model. In order for the system to be Amson invariant, the constraint must be applied separately for each particle pair interaction. For example, in a three-particle system where (1) and (2) are matter and (3) is antimatter, we have the constraints

$$\overset{(1)}{m}_a = \begin{cases} +\overset{(1)}{m}_i & \text{for } \overset{(2)}{m} \\ -\overset{(1)}{m}_i & \text{for } \overset{(3)}{m} \end{cases} \tag{16}$$

$$\overset{(2)}{m}_a = \begin{cases} +\overset{(2)}{m}_i & \text{for } \overset{(1)}{m} \\ -\overset{(2)}{m}_i & \text{for } \overset{(3)}{m} \end{cases} \tag{17}$$

$$\overset{(3)}{m}_a = -\overset{(3)}{m}_i \, . \tag{18}$$

While the magnitude of m_a depends only on the particle in question, the sign of m_a depends on the relative character of the interacting particle. Active mass, then, is not an inherent property a particle but rather is determined by particle interaction. Our only justification for the dual nature of m_a is Amson invariance. Although nothing in Newtonian mechanics forbids this, it lends no support either. For that, we must look to general relativity.

V. MODIFIED EINSTEINIAN GRAVITY

As we have seen, crossing symmetry does not preclude the strong equivalence principle, which is central to Einsteinian gravity. In fact, if one considers a purely matter or antimatter universe, then general relativity may be applied directly, without modification. However, in an Amson invariant universe in which both are present, the geometric description breaks down. It is the active mass which defines the metric of a space-time, and therefore one cannot derive a global metric due to the dual nature of m_a.[6] By imposing crossing symmetry, we would seem to have entered an intractable forest. We are prevented from describing gravity as a gauge-invariant field, and we are forbidden to geometrize the gravity in a consistent manner. However, recent work in discrete gravity may provide a path out of the woods.

Souza, and Silveira [14] have developed a discrete and finite gravitational model by invoking extended causality rather than the usual local causality. In standard field theory, causality constrains the evolution of a field by

$$\Delta\tau^2 + \Delta\mathbf{x}^2 = 0 \tag{19}$$

where $\Delta\mathbf{x}$ is a four-vector in the locally flat space-time. This defines the light-cone of a point in the field. Extended causality is more restrictive, in that it applies causality simultaneously to the region about a point. This requires in addition to (19),

$$\Delta\tau + \mathbf{f} \cdot \Delta\mathbf{x} = 0 \tag{20}$$

where $f^\mu \equiv dx^\mu/d\tau$ is a constant, space-like, four-vector tangent to the light-cone: a cone generator.

---

[6]Since Amson invariance does not violate the strong equivalence priniple, one may still derive a local metric at every point in the space. For this reason, Amson invariant gravity is a metric theory, despite the lack of a global metric.

This additional constraint limits the form a field can take. In standard field theory, a field $A(\mathbf{x}, \tau)$ is defined on the three-dimensional light-cone, whereas in the field theory of extended causality, a field $A_f(\mathbf{x}, \tau)$ is defined only on the straight line of $\mathbf{f}$. The A_f field represents a single real quantum of the field, propagating along $\mathbf{f}$, while the A field corresponds to a continuous distribution of imaginary field quanta. In other words, the field exists *only along* $\mathbf{f}$. The global field A is merely a construct which aids in visualization and has *no physical reality.* Furthermore, A_f is only generated by an interaction between two particles. If there is no interaction, there is no field.

Consider then the gravitational field of a central mass. In traditional general relativity, the gravitational field is simply the Schwarzchild metric. It is assumed that the field exists independently of any test masses. However in the discrete gravity approach, a single mass has no gravitational field, thus its metric may not be derived. It is only when a test mass is introduced that the field exists, and then only along the line connecting the two masses. Thus, a test mass would detect a Schwarzchild metric wherever it is placed, and one might infer that the Schwarzchild metric exists everywhere. In actuality, the space-time is flat except along the straight line connecting the two masses. The metric of the central mass is completely flat in the absence of all other masses. In general relativity, the gravitational field is generated by the active mass, m_a. In the discrete gravitational model, the gravitational field is generated only by interactions between particles. It would seem then that discrete gravity requires m_a to be generated by particle interaction.

VI. CONCLUSION

There is still much work to be done in this area. In particular, the Souza-Silveira model makes discrete a continuum field model, while the Noyes-Starson model is *a priori* discrete. It is not certain that the two are compatible, though the likelihood is good. It is clear, however, that matter-antimatter repulsion does not contradict current theoretical models outright. Rather it relegates them to special cases of a much richer universe.

---

[1] see, e.g., D B Lichtenberg, *Meson and Baryon Spectroscopy*, Springer-Verlag (1965).
[2] C Wu, et al., *Phys. Rev.* **105**, 1413 (1957).
[3] S Weinberg, Phys. Rev B **138** 988 (1965).
[4] J Amson, "BI-OROBOUROS — A Recursive Hierarchy Construction" Proc. ANPA 7 (1985).
[5] H P Noyes, "From Bit-Strings (part way) to Quaternions" SLAC-PUB-5431 (1991).
[6] H P Noyes and S Starson, "Discrete Anti-gravity" Proc. ANPA WEST 7, F. Young, ed. (1991).
[7] A Einstein, "The Foundation of the General Theory of Relativity", in *The Principle of Relativity* Dover (1952).
[8] C W Misner, K S Thorne and J A Wheeler *Gravitation* p189 Freeman (1973).
[9] C Moller, *The Theory of Relativity* p220 Clarendon (1952).
[10] R v Eötvös, *Ann. d. Phys.* **68** 11 (1922).
[11] B S DeWitt and R W Brehme, *Ann. Phys.* **9** 220 (1960).
[12] A Papapetrou, *Proc. R. Soc. London A* **209** 248 (1951).
[13] H Bondi, *Rev. Mod. Phys.* **29** 423 (1957).
[14] M M Souza and R N Silveira, "Discrete and Finite General Relativity" gr-qc/9801040 (1998).

An Extension to Eddington's Fundamental Theory *or* Eddington Off Limits

Keith G. Bowden and Clive W. Kilmister

Reprinted from Proceedings of ANPAWest, Stanford, February 1995

The train of thought that led, eventually in 1962, to the Combinatorial Hierarchy was started in the 1930's by Sir Arthur Eddington. The numerical skeleton of ANPA's derivation of the coupling constants runs something like this. The coupling constants of (grand) unified field theory appear in the Hierarchy as the number of DCS's at each level of the Hierarchy. These are given by the **cumulative sums of** the recursion

$a(i+1) = 2^{a(i)}-1$ with $a(1)=3$,

provided $a(i)<b(i)$ with $b(i)$ defined by the recursion

$b(i+1) = 2^{b(i)}$ with $b(1)=4$.

Note that $a(0) = b(0) = 2$. An important part of the argument for the ANPA contention was the cutoff of the recursion, caused by the **b** series growing more slowly. This was thought to mark its superiority over Eddington's approach. The following is a neat bit of Common Lisp, coded by Alex Bowden, and its output. It gives, before the computer overflows, the number of DCS's in the four levels of the Combinatorial Hierarchy.

```
>>>
(define (anpa n)
  (do ((i 1 (1+ i))
       (tl 0 (+ ai tl))
       (ai 3 (- (expt 2 ai) 1)))
      ((= i n))
      (display (+ tl ai))
      (newline) (newline)))

>>> (anpa 5)
```

3

10

137

170141183460469231731687303715884105864

```
ERROR:  Numeric overflow in *

/*        Well        it's        got        about
170141183460469231731687303715884105864  bits  so  it's
not going in 12 meg of memory, is it? (Not quite the
same kind of cutoff as Parker-Rhodes discovered but it
has much the same effect.) */
```

It is ANPA's claim that the last two numbers in this series are approximations to the Fine Structure Constant and the Gravitational coupling constant, and that the first two are associated with the weak and strong nuclear interactions.

Some time ago I was prompted to look at Eddington's original derivation of the Fine Structure Constant. His logic, as I understand it, goes something like this. The automorphism from 3+1 space-time into itself is a 4 by 4 tensor mapping with nine (=3×3) space-space like and one time-time like element (seen in Fig (1) as block matrices along the leading diagonal of the mapping tensor) giving ten homogeneous and six remaining heterogeneous elements.

Fig*(1)*
$$\begin{bmatrix} x & x & x & . \\ x & x & x & . \\ x & x & x & . \\ . & . & . & x \end{bmatrix}$$

This logic can be continued in the same vein. The automorphism from the space so constructed into itself is a 16 by 16 tensor mapping with one hundred (=10×10) homogeneous-homogeneous and thirty six (=6×6) heterogeneous-heterogeneous elements forming the 136 elements in the leading diagonal of the block matrix in Fig (2). It is Eddington's claim that these 136 connections give a first approximation to the Fine Structure Constant. His argumentation was obscure, a fact which was not helped when a more accurate determination of the empirical value of α led Sir Arthur to add an arbitrary correction of 1 to this value..

Fig(2)

$$\begin{bmatrix}
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
x & x & x & x & x & x & x & x & x & x & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & x & x & x & x & x & x \\
. & . & . & . & . & . & . & . & . & . & x & x & x & x & x & x \\
. & . & . & . & . & . & . & . & . & . & x & x & x & x & x & x \\
. & . & . & . & . & . & . & . & . & . & x & x & x & x & x & x \\
. & . & . & . & . & . & . & . & . & . & x & x & x & x & x & x \\
. & . & . & . & . & . & . & . & . & . & x & x & x & x & x & x
\end{bmatrix}$$

Out of curiousity one day I wondered why Eddington had not extended this train of thought in the obvious way. (Perhaps because he had not got access to a hand calculator?) It was a source of some embarrasment to him that higher numbers did not seem to be needed, because he believed that everything in a (correct) theory should be realisable. He was forced to a weak argument for putting in limits, viz "Physics does not seem to go to a more complex analysis." Little thought is necessary to show that the recursion (**not cumulative** this time!)

$a(i+1) = a(i)^2 + (b(i) - a(i))^2$ with $a(1)=3$ (number of space-like elements),

with b(i) as before defined by

$b(i+1) = 2^{b(i)}$ with $b(1)=4$ (number of space-time elements).

does the trick. Again the code is by Alex,

```
>>>
(define (eddington n)
  (do ((i 1 (1+ i))
       (ai 3 (+ (expt ai 2) (expt (- (expt 2 (expt 2
i)) ai) 2))))
      ((= i n))
      (display ai)
      (newline) (newline)))
>>> (eddington 13)
```

```
3 /* 1.  As ANPA.*/

10 /* 2. As ANPA. */

136 /* 3. Eddington's first approximation to the Fine
Structure constant */

32896 /* 4. What is this? */

2147516416 /* 5. And this? */

9223372039002259456 /* 6. And this? */

170141183460469231740910675752738881536    /*    7.    An
approximation to the Gravitational constant! */
```

Looking at the results of this program we find that if Eddington had extended his argument to the seventh term he would have ended up with an approximation to the Gravitational constant which is as good as ANPA's, agreeing with it to one part in 10^{20}! Of course there is some analytical reason for this agreement. It is easy to prove that $\lim_{i\to\infty} a(i)/b(i)=1/2$ which shows this number to be near $2^{128}/2$. But the closeness of agreement *is* a surprise. Two other facts are quite stunning. Firstly, that the first two terms of the Eddington series turn out to have the same values as the first two terms of the ANPA series, but more importantly that we now have three new numbers that remain to be interpreted. The fact that two such (interpretable!) number series exist, each containing the sequence 3, 10, ~137, ~1.7×10^{38} is remarkable in itself. Perhaps ANPA should reconsider the difficulties with interpreting the first two terms in the Hierarchy series in terms of Eddington's .

For the record the next few terms of the Eddington series are

```
57896044618658097711785492504343953926805133516280751 25
14604793076724489256 96 /* 8. Etc...*/
```

```
6703903964971298549787012499102923063739682910296196688
8617807218608820150368313844455558071811634993375202730
47169660419209761687993013765220781067862016 /* 9. */

8988465674311579538646525953945123668089884894711532863
6715040578866337902750481566354238661203768010560056939
9356966788293948844072083112464237153197370688927879116
8373129242516360890354611079940945177223919174619979720
2053246477510444956207562702033389100358222242362999049
846181264125702651381636615110656 /* 10. */

1615850303565550365035743834433497598022205133485774201
6065172713762327569433945446598600705761456731844358980
4609490097470597795752454605475440761932241415603154386
8365049804587509887519482605339802881919203378413839610
9321309878080919047169238085235290822926018152521443787
9457705329043037761995619651927610470513515772870057289
5265282173598410898686661027635767561787024437943425698
0638641483535093292326373531858799166625357628570460558
0312634482725219363806778024467890678428475007296606640
88708794565812 /* 11. */

5221944407065762533458763553583121912899821245236918901
9211674164197695398577872842441340596749877917044505335
7219631418993786719092896803631618043925682638972978488
2718549991701807950671918591572140350059279731131881594
1969885637283616734217229330874840395435290185203564202
4370059304557233988891799014503343469488440893892973452
8150951304702997897267164117346515133482215295125079861
99933857107770846917779942645743159111895721724836704390
5936319748237550094520674504208530837546834166925275516
4860441347753845213633709579064526157457936026174623807
5035455059100025373721570856630568171033987937658200940
8041735518192776370118226857395422706084137899283135311
0840931347218060378305010831919411921868028097573137521
0496 /* 12. Penrose */

5453740678097079647314921223668914312241320809981163462
1591639309486066592455964760813211726260099361197864589
8078512636554935410088592031805489882538777399539453149
4210964947693049126140241025798484258067958190983859432
7130466228006064527695094315050895012626789995860000503
9800013267918400452648902940476175250815097737826955502
```

```
6561822800074237130176467756229219644593763843481396720
4402780875784717497270333891257040745030805296012821925
2289006663246782918023621203691221406122565758878759582
4496131828718624188187815726380568132931419168431057899
6819010439268837772668394957847117216977833157535043606
7677351278351560020653627479172541787198269144680385404
8927528945648395367639002746781078054539792258647705798
6464262133204575856257447935634823314580164346005450598
6454390370325727762800915817085785916089384093740274639
8772032937216835758854590104623171902925044242535836564
0500647564707761686935606296723808982721408215159160802
5112310870383454816022453015044863226577398727115457604
5695777415368938864187796833086518025304965833044106946
9963255718232392027772110055656904417888060344991024842
2007733476867745360207712406657576020170089071399056726
9908318266418984495517241000628097353273428028435760453
4538151543976088875005983057931660359796131876092811047
7088915166014137064241943234762333647731605028190440195
0693965309980296254796623905849388480668426224768340708
4966393100366463003468636791034920118761204956070029394
34991616 /* 13. Penrose likes big numbers like this! */
```

References

1. Eddington A. S. [1933], *The Nature of the Physical World.* Cambridge.

2. Eddington A. S. [1934], *New Pathways in Science*. Cambridge.

Παντα ρει

An Interview with Prof. Basil Hiley

Birkbeck College, March 11, 1999

by Arleta Griffor

Arleta Griffor: Do you remember the conference, 'Quantum Theory and Beyond', (Cambridge 1968), organised by Ted Bastin and David Bohm?

Basil Hiley: Yes, I still have a vivid memory of it. It was a rather interesting conference in the sense that it was trying to get away from formal presentations. Bohm's constant theme was that he did not like physics conferences where you would get up give a paper, sit down, and then someone else would get up and give his paper, and there would be no reference to each others papers. The idea behind the 'Quantum Theory and Beyond' conference was to go with no prepared conference papers, but to discuss various ideas as they arose during the meeting. Of course people did prepare something but there was much more effort to have a meaningful discussion. In this way it was a very interesting experience.

There were two discussions that stood out for me. One was Penrose's discussion of spin networks, but that was part of an ongoing debate that I had with him at Birkbeck College. It was nice to hear him present his spin-networks at one session rather than picking up bits and pieces of it as in our discussions. The other talk, which was actually more interesting for me was one given by Ron Atkin from Essex University. He also gave me an extremely interesting tutorial on cohomology theory while we were walking around the outside of King's College Chapel. I remember that afterwards we got into a little bit of argument because he felt I was pinching his ideas from his paper when I wrote my talk after the meeting. I was certainly helped enormously by him. But Bohm and I had already been developing the essential ideas and were struggling to provide the correct mathematical structure. I was able to do this thanks to Atkin's help.

A.G. The discussion following Bohm's conference talk indicates that Bohm was already considering a new concept of order, more fundamental than space-time structure. In the discussion, C.F. von Weizsäcker and Ted Bastin were trying to get out of Bohm what he meant by his concept of 'order'. What was your position at that time?

B.H. I was more or less on the side of von Weizsäcker and Bastin in trying to understand what Bohm was actually getting at. Although I had been discussing these ideas of order with him, I was struggling to find my own ways of thinking about these things. I came into this problem as someone who had been steeped in conventional physics, and had not thought about these deeper problems. A lot of these arguments just went over my head at that time. This was confirmed later as I read the Biederman-Bohm correspondence[1]. In those letters I saw the idea of the implicate order emerging, but at that early stage I was not fully aware of what he was getting at when we were having those discussions about order.

A.G. What were you working on at that time?

B.H. I remember that we held seminars every Thursday afternoon in the physics lecture theatre at Birkbeck College. It was mainly David Bohm talking about his ideas. At times we were joined by Penrose and by Erwin Kronheimer who was a topologist in the Mathematics Department. We were talking about the structure of space-time; about the possible role of combinatorial topology and cohomology in our general discussions about order. Bohm went through Lefschetz's book *Topology*. He would come in each Thursday with the book in his hand and would go through it page by page presenting the arguments. Every so often he would put the book down and begin to talk about how these ideas would fit into his general scheme so that we could fit quantum mechanics into a topological framework. We also went through Hodge's book, *The Theory and Applications of Harmonic Integrals*. We were doing this before I met Atkin so when I talked with him at Cambridge a number of difficulties that I had had disappeared! Indeed, I put these ideas into my paper in Bastin's book *Quantum theory and Beyond*. That paper was my very first adventure into this particular area.

A.G. So, are you saying that already at that time you were working on the idea of order?

B.H. Yes, our investigation of combinatorial topology was just that. We were trying to get away from the continuum and from the Cartesian notion of co-ordinates. What we were trying to exploit was the cohomology defined by $\int pdq$ around a circuit, which in quantum mechanics is an integer. If we exploit the isomorphism between de Rham cohomology and an abstract cohomology based on chains and co-chains, we can provide a structural description of quantum mechanics without referring to the continuous nature of space-time. The integral can be expressed in terms of the intersection of a chain and the corresponding co-

[1] *Bohm-Biederman Correspondence, Volume One: Creativity and Science*, ed. P. Pylkkänen, Routledge 1999

chain, which must be mapped on to the integers. This was one very simple idea that we were trying to exploit. Our proposal was that as the quantum mechanical properties did not seem to depend upon the local properties of space and time, but were sensitive to global conditions, i.e. the quantisation condition, we could use a discrete topological description to capture the essence of quantum processes. We also noticed that all the main equations of physics could be written in a very simple topological form in terms of boundary and co-boundary operators. Was this a trivial 'discovery' or was there something more in it? Together with Stuart, Bohm and I wrote the paper entitled *On a New Mode of Description in Physics*. In this paper we discussed all of these ideas, but at this stage we were essentially talking about a discrete structure that does not require space-time to be 'there'.The description is in terms of abstract simplicial complexes using boundary and co-boundary operators to analyse the structure.

A.G. Did you have any contact with ANPA people then?

B.H. There was some exchange. I remember Clive Kilmister organized a series of seminars at King's College on the nature of space and time. Clive talked about space-time from the relativistic point of view, I talked about space and time from the kind of things we were doing at Birkbeck. Richard Sorabji told us about the Greek views of space and time. It was there, after giving my talk, that a young American research student asked me: 'Have you ever read Grassmann's book *Die Ausdehnungslehre*? I said: 'No, because I don't read German too well'. He immediately replied that you don't have to read German, because a guy had translated some of Grassmann's ideas into English. I did have some experience of Grassmann because I had read a book by Forder, *The Calculus of Extension*, which was in English, and that was extraordinarily interesting. So when this American student said: 'Go and read Grassmann', I did just that and found it fascinating. I discovered that Grassmann was actually talking about thought, of how one thought becomes another: he was talking about 'becoming' not 'being'. It was these ideas that led him to what we now call "Grassmann algebra". Algebra was a way to talk about becoming!

A.G. Was that the beginning of your algebraic approach to quantum theory?

B.H. Yes, that was a decisive moment in the work for me. What I found fascinating was that some of the topological notions like boundary and co-boundary operators could be formulated in terms of aggregates in a Clifford algebra. Bohm and I wrote about this in what I regard as an important paper entitled *On the Relativistic Generalisation of Phase-Space*. Here we presented these ideas which fused together topology and algebras. This was a stepping stone, the first paper indicating where we were going.

In the background, however, there were so many different investigations going on in parallel; the light-cone structure was being used to generate space-time. An example of this was the Penrose spin network that he was trying to generalise using twistor theory. We have been very interested in this theory, but we were investigating twistors from the point of view of the conformal Clifford algebra, which was not the way Penrose was going. A lot of that work was never written up, but some of it appeared in a paper written by Bohm and myself entitled *Generalisation of the Twistor to Clifford Algebras as a basis for Geometry.* Unfortunately we published it in Revista Brasileira de Fisica! At that time we were considering variety of different ways of looking at how we might be able to do physics without actually using the space-time continuum.

A.G. Why didn't you want to use the space-time continuum?

B.H. If you read my paper in *Quantum Theory and Beyond,* you will see that I was using the idea of a cellular phase space. The reason for this was to focus on the uncertainty relationship, $\Delta x \Delta p$, which is an invariant in quantum mechanics. So what I was doing was taking the uncertainty relationship as basic and building from there. Thus any system can only be localised in some cell in phase space. One could then build up a structure of these cells, and that was what I was doing in that paper. So I wanted to start from the uncertainty principle and build it in to my structure right from the beginning. In this way I was building the uncertainty relation into the description, not as Bohr did, but through topological structures. Thus the basic order in quantum mechanics was to be based on discrete structures.

There were other reasons which encouraged me in this direction. Bilby and his group were investigating the dynamics of crystal dislocations. What they found was that these crystal structures gave non-Riemannian geometry in the continuum limit. Thus a discrete structure could give you interesting continuum limits that you might be able to exploit. I have already remarked that $\int pdq$ around a circuit was an integer multiple of $\hbar$ We could then regard this condition as the appearance of a series of edge dislocations in the structure.

In the case of quantum mechanics we could form this structure from waves. The momentum p could be thought of as a describing a set of planes as in Hamilton-Jacobi theory. The Δq's could be thought of as elements of a curve. When we integrate around a circuit we count up the number of planes crossed by the circuit. By suitable choice of orientation, we can count the number of edge dislocations enclosed by the circuit. Furthermore the value of $\int pdq$ around a circuit is independent of the exact shape of the circuit; it only has to enclose the singularity.

Remember quantum numbers classify the stable states of an atom and conventional quantum mechanics does not enable us to understand these states in terms of electron trajectories. Thus the quantum numbers are insensitive to the path of the electron, as is the circuit integral. Can we therefore associate the quantum numbers with dislocations in the structure of space?

Another reason for exploring discrete space-times comes from the need to avoid the infinities we get in field theory. If you could do a field theory in a discrete space, you could avoid these infinities. What we found was that the people working with crystals had already shown that the Biot-Savart law of electromagnetism could be explained in terms of dislocation theory. Here the electromagnetic field appears as crystal stresses around dislocations. You can also show that as a dislocation moves in the crystal, it experiences effects similar to those found in relativity where the speed of sound plays the role of the speed of light. So all these analogies that were floating around spurred us on with the investigation.

A.G. As I understand it, these were your reasons why you wanted to get away from the space-time continuum.

B.H. I am describing how I saw all this work. Certainly Bohm encouraged me to go down this line, but the main drive did not come from him, it came from me. He was developing what became the implicate order, but that term did not crop up in our discussions as far as I remember. I did not really become aware of the significance of the implicate order until the 1980's. I remember when I read Bohm's paper *Quantum Theory as an indication of a new order in physics*, I was somewhat puzzled by the notion. The statement "What the hell is this?" probably best summarises my reaction. I did not realise this was all there in the background while we were talking together.

I must also mention Bohm's inaugural lecture, which was written up as *Problems in the Basic Concepts of Physics*. It was a beautiful paper. Bohm was aware of all of these discussions that were going on, but his ideas were so general that it was difficult to put a mathematical handle on them. We had these topological ideas, but I felt that everything was too static, nothing was changing. It was all very disappointing because we wanted ultimately to base our work on process, or movement as we called it at that time. Bohm was always talking about 'movement': 'Movement is what is', he would say. How can we capture movement in our mathematics? How do the artists do it? How do they capture movement? I went to the Tate Gallery looking for inspiration, but didn't find any! Incidentally the recent publication of the Bohm-Biederman Correspondence reminds me that I actually went to see Biederman's work at the Tate.

A.G. What did you see in these pictures?

B.H. Absolutely nothing, because I did not understand what Biederman was getting at. There were a few planks of wood crossing each other, but so what? What I was much more interested in was how the painters captured movement in things like waterfalls. David was trying to suggest that there was some connection between the mathematics of Hilbert space and the way Impressionists tried to achieve their effects.

Bohm was trying to make me think in a new way. It was to do with the importance of relationship. When you begin to paint something, the choice of what colour to use in one particular patch of the canvas depends on what you are going to put in another region not only locally, but also at other patches that were some distance away. There was a kind of wholeness and the relationships within this totality created movement.

A.G. Context dependence.

B.H. We did not call it 'context-dependence' at the time. I don't think we ever used the phrase 'context dependent'. Perhaps we should have done, but I don't remember this being a significant phrase. That all emerged later. At that time it was more a question of articulating a structure of relationships. We start with relationship, then the relationship of relationships that form a structure, and finally end with order. So the most primitive thing, for me then, was relationship.

In the case of the patches of paint, what you are doing is looking at the relationships between the patches of paint in different regions. So the overall picture depends not only on this one region here, but also on its relationships to its neighbours and even further afield. In other words we were trying to discuss order through relationship and structure. That is where the simplical complexes came in. They were the basic building blocks of our structure.

In this way we obtained structure, but no movement and we needed movement. Remember, we called it 'structure-process'. Process was missing. All we were getting when we were doing the mathematics was a frozen structure with nothing moving. However the artists had captured movement. There is a beautiful picture of a waterfall painted by Ruisdael in the National Gallery. The water actually looks as if it was moving. Whenever I have tried to photograph a waterfall, it did not look like a waterfall at all, it was flat and static, nothing seemed to move. But if you see Ruisdael's picture, as you look at it, you can actually see the water moving. Ruisdael had created movement through a structure of relationships. My question was, can we create the idea of movement from within mathematics?

A.G. And how was the Grassmann algebra relevant here?

B.H. It wasn't so much the Grassmann algebra as such, but what Grassmann was trying to achieve. As I remarked before, Grassmann argued that mathematics is primarily about thought, and not about material process. This idea was vital for me because as a physicist I was steeped in the importance of mathematics as a tool for studying material process, in solid state physics, in electromagnetic theory and all the rest of it. I was taught that mathematics was about the material process and I believed it! I thought that mathematicians take the ideas that physicists have generated, together with their simplistic attempts at mathematisation and then made it all rigorous and more general. That was the way I was thinking at the time. Then, when Grassmann came along, for whom I had tremendous respect, I saw that for him mathematics was about thought, about becoming, about process, and its use to describe material process was secondary. He turned my cosy physicist's world upside down. Once I realised that mathematics was a description of thought, I found I could be free from all the traps that I was in as a physicist. That is why I keep stressing this point in my papers. I found it very liberating.

I did my PhD on crystal lattices where everything was static, nothing moved. However now the idea was that we are not just putting together static things, but we were trying to put together things that were moving, changing. Recall Bohm's stock phrase: *movement is what is.*

A.G. How exactly did you represent movement in terms of your algebraic language, and what kind of movement was it?

B.H. This is a difficult question to answer directly. It didn't come easily, as a linear progression of ideas. It appeared as an impressionist painting takes shape, with little 'patches of ideas' being pulled together. In the background was Bohm's structure-process. In quantum mechanics there were the matrix elements representing transitions (transition probability amplitudes). There was the realisation that the wave function was a transition probability amplitude, not a 'state' function in the literal sense. There was the realisation that the spinor 'wave function' was an element of the orthogonal Clifford algebra; that the ordinary wave function was an element of the symplectic Clifford algebra. Everything in quantum mechanics could be described in an algebra. Could this be the root of process or movement?

Then came Grassmann's ideas. What lay behind the Grassmann algebra was 'becoming', not the static geometric structures that we see in the books today. All traces of the process has been completely frozen out. Schönberg had shown us how Clifford algebras could be constructed from a pair of Grassmann algebras. When I looked into how Clifford arrived at his algebras I found he was talking about 'rotors', and 'motors', about movement. Thus at the heart of his thinking was this idea of movement, of activity, of change. It was Gibbs, at the turn of the century, who said that we didn't need things like the Clifford aggregates, the 'rotors' and 'motors' because the multiplication of the elements changed vectors into bivectors, bivectors into trivectors and so on. The geometric entities like vectors, bivectors, trivectors, etc., did not transform into each other, vectors transformed into vectors and nothing else and so on. So Gibbs argued that we didn't need Clifford's ideas in physics; we needed geometric description.

When Dirac discovered that the electron needed the spinor, a minimal left ideal that naturally arises in the Clifford algebra, no one found it necessary to look at what Clifford had been discussing in his original papers. In fact he complained that people had missed the point by giving preference to static images rather than recognise the structure that lay behind his rotors, motors, and so on.

It slowly dawned on me that algebra was the vehicle I needed to carry out the exploration of a process-based approach. Furthermore Grassmann had laid the foundations of this approach, but his thinking did not permeate through to the textbooks. What I was able to do then was to develop my views on the algebra of process. We have a way of actually capturing the essence of movement or activity in terms of the elements of an algebra. Each movement has a beginning and an end (Hegel's Sein und Nichts). Rather than representing the beginning, a, and the end, b, separately, we use the Grassmann symbols [ab], [bc] and so on. The bracket signifies that the beginning and the end cannot be thought of as separate entities.

It was the 'flow' or process that we wanted to capture. Then, by making simple rules of combination of these symbols, I could generate the orthogonal Clifford algebra. In other words, I could discuss rotations by thinking about movement from a to b, movement from b to c, and then the movement from c to a. Putting simple rules on the combinations of these movements produced the Clifford group. I seemed to be able to carry the rotational symmetry, without the need to set up co-ordinate systems or even refer to an underlying space. This reminded me a little of what Penrose was trying to do, but he was not doing it in terms of movement, he was doing it in terms of spin-networks. When you statistically average over the Penrose spin network you get the orthogonal group coming out. I didn't need to average, but obtained the orthogonal group without even mentioning space or time at all. I just got it out of movements.

A.G. Are you saying that the elements of this algebra, such as [ab], [bcd], etc. are movements, or represent movements?

B.H. Yes, exactly! They are all different forms of movement. This idea came from Grassmann. As I said before, Grassmann raised the question about thought. His question was: 'Is the new thought independent of the old thought?' It is quite clear that it is not, because thought is about becoming. Any thought has the potentiality of a new thought already in it and the new thought has the trace of the old thought in it. Therefore we should really consider both 'sides, or both 'ends' together. Hence the symbol [ab] used to describe the process of becoming.

A.G. So the elements which represent movement consist of at least of two 'ends'.

B.H. Yes, at least two 'ends', but the word 'end' could be misleading. I prefer to call them 'poles'. Here the analogy with magnetic poles is close. Recall you can never separate the north pole from the south by sub-division. You always end up with two poles.

In a more complex process one could imagine having more poles (quadrupoles for example). One can have three or four, and so on, but you start with two. I like to think of this algebra as defining a directional calculus. In other words, you are starting at some region, and by using these movements - I have got no space to extend the movement into yet - but just by using pure movements, the mathematics carries the rotational symmetries.

The next thing was to relate different regions. This could be done by using the Penrose twistor. In this approach you could construct the light-cone at a point using the Dirac spinor. That would be similar to using the Dirac Clifford algebra to set up a directional calculus. The twistor can then be used to relate different light-cones at different points. I could do a similar thing by using the conformal Clifford algebra. So what I could get from a very simple set of rules was the content of the twistor algebra in terms of the structure of movements.

I have not written these ideas up in the form of a paper, but I have a set of lecture notes setting this all down. I should have typed it up but I have never got round to it! It was all going into a book that I planned to write with Fabio Frescura on Clifford algebras. Unfortunately Fabio was forced to return to South Africa and the book never materialised.

Another reason why this never got written up was because I was not happy with the twistor idea. It did not have the right kinematics in it. I felt this would be provided by the symplectic Clifford algebra. The mathematical structure of this algebra really stumped me for a very long time. I tried looking in the mathematical literature for help, but there was very little that I could find that was useful. The orthogonal Clifford algebra is very easy to deal with, it is all finite. There are a finite number of elements and you have finite dimensional representations. All you need is to know how to deal with matrices. The Dirac algebra has sixteen elements, and the generators can be represented by four-by-four matrices, and that is it! The symplectic algebra is an infinite algebra, and physicists have more difficulty dealing with an infinite number of dimensions! However in the end we got there and this work is now being written up.

In the meantime, Bohm told me that I ought to look at one of the algebras that Weyl discusses in his book *The Theory of Groups and Quantum Mechanics*. The algebra was finite and became the symplectic Clifford in the infinite limit. He did not tell me what to do with it, but just that it might help in understanding the infinite algebra. It was while looking at this algebra that I came up with the idea of how I could define 'points' in my algebra, which in the limit would become like the point of the continuum. So the algebra had not only the movement structure in it, but it also had a point-like structure. This point was not static but could be understood as being a movement that continually moves into itself. The process P follows itself P, and remains itself P, that is it is idempotent $P^2 = P$. In this way I was able to 'create' a discrete space containing a finite number of 'points'. In fact in this way I constructed a discrete phase space so that it contained kinematics and furthermore that kinematics carried a finite version of the Heisenberg algebra.

A.G. To make it clear, you use the word 'point', but you understand your 'points' in a different way from the way they are ordinarily understood, that is, as points of space-time continuum.

B.H. Yes, I am looking at them very differently. They are embedded in the algebra; they are part of the algebra. These points can be labelled using the Dirac position kets |x⟩. They are generalised points that can be regarded as 'quantum points'. This is the interesting feature.

A.G. So you say that the apparently structureless points of space-time have this algebraic 'inner structure', as it were. Is your algebra carrying all the transformations that are necessary to define space-time?

B.H. Yes, but time presents more difficulties. I have discussed time in my paper with Marco Fernandes *Process and Time*, but as far as space is concerned I do not see any problems.

A.G. In Clive Kilmister's note *ANPA and Birkbeck - Connections and Differences*, he says that your algebraic approach begins 'some way down the line', or at what he calls a higher level, comparing it with his own approach.

B.H. I don't disagree with this position at all. As I keep telling him : 'Clive, if I knew how to start from the first level, I would start from the first level'. What I am doing is exploring the algebras that I already know have significance in physics.

A.G. And what is the physical significance, or physical meaning of your algebra?

B.H. I have in the background the implicate order. To illustrate what I have in mind start with some general explicate order t_1. Now the idea is that order is enfolding itself back into the implicate order and then unfolding again, giving a new explicate order t_2. Thus t_1 and t_2 are just two explications of the general unfolding process that is continually on going. These two orders are being made manifest in succession. At this stage they are quite general because I am speaking very generally, but these are the images that I am using. They are my images of what I mean by an unfolding process or movement.

A.G. In some of his writing, Bohm says that the whole algebra can be understood as the implicate order, or a holomovement, and certain structures of the algebra can be understood as explicate or implicate, and certain operations as enfoldment, and others as unfoldment.

B.H. This is exactly what I am trying to get over. I have the Weyl algebra which represents a 'toy' implicate order. In it I have structures that have the properties of generalised points. What I have not stressed so far is that I also have momentum-like points in my algebra. To make these explicate I must make an inner automorphism on the algebra. This will enable me to display the generalised momentum points. But to do this my position points are folded back into the algebra. Thus this algebra describes the implicate order in which both momentum space and position space are implicit in it, but only one set can be made explicit at any one time. This is how the notion of complementarity fits in. When we started, you asked if it was merely space, but remember, I answered: 'No, it is a phase space'. And this phase space has the quantum character built into it right from the start. I don't start with a classical phase space and then try to quantize it.

A.G. I am puzzled that these t_1 and t_2 are two different explicate orders. Where in your algebra is the enfolding-unfolding movement? How it is represented?

B.H. Lets go more deeply into the unfolding process. We have called the first explicate order t_1. This is followed by the enfolding M_1 which takes t_1 into an implicate order. Thus the outward movement is M_1. The next inward movement is represented by M_2. This process is written as, $M_1t_1M_2$. We then assume this process gives the new explicate order so we write $t_2 = M_1t_1M_2$. In other words this new explicate order is just the combined process. I know this is not what I would call a compelling argument, but let's suppose that folding and unfolding are similar processes in which no new content has been introduced. We could then write $t_2 = Mt_1M^{-1}$. We can simplify a little bit more and write M as $e^{-i\hbar\tau}$. Here we parameterise the folding by some parameter τ, then what we end up with is Heisenberg's equations of motion.

A.G. Are all t's and M's elements of the same algebra?

B.H. They are the elements of the same algebra, but remember the explicate order is a distinguished order in the implicate order. I made a distinction by calling t the explicate order, that is the order that can be made manifest. The M can be regarded as the subtle movement.

A.G. So this inner automorphism represents enfolding.

B.H. Yes, it represents enfolding. So the inner automorphism on the algebra always represents enfolding and unfolding.

A.G. Why only the inner automorphism and not outer? Is there any special significance to enfolding being the inner automorphism?

B.H. You can use outer automorphisms as well, but you would have to give them some specific meaning. I would not object to exploring different outer automorphisms. All I have needed to date are inner automorphisms.

A.G. Does discreteness have any significance here?

B.H. No, we started with discreteness but we don't need it; it is not a necessity. For example, the Weyl algebra I use is discrete, but you can always go to the continuous limit. Bohm argued that it was not a question of a theory being either discrete or continuous. We actually need both! At one level it could be discrete, while at another level it could be continuous. For example, the Weyl algebra could be made continuous, but if you look at the eigenvalues of its elements, they can be discrete, therefore you are carrying discreteness in the continuum.

A.G. Why I ask about that is because for ANPA people discreteness is very important. I suppose it is because of their concept of process as a structure of information, and information is thought to be discrete.

B.H. They treat it as epistemology, whereas I was trying to approach things from the ontological point of view.

A.G. This is one difference.

B.H. A big difference.

A.G. The other difference is, as I said, that they treat process as a structure of information. They say, 'process is the process of increase of information'. Is it not so that information has always to be discrete?

B.H. This is an interesting question, but I don't think information has to be discrete. For example we have argued that in the Schrödinger picture we can abstract the quantum potential and as we have explained in *The Undivided Universe*, the quantum potential carries active information. But the Schrödinger representation is isomorphic to the Heisenberg algebra for finite degrees of freedom. This means that there must be a role for information in the Heisenberg algebra. In other words, the algebras already contain an aspect of information.

A.G. But whenever the concept of information was considered, it was thought about in terms of bits, and any structure of bits is discrete.

B.H. I think this arises because we are using the information in a different way. The basic problem is to find out exactly what we mean by the term 'information'.

A.G. Bohm understood information in the sense of 'in-form', 'to form from within'. Would you say that his concept of information was different from the concept of Shannon information, and therefore, it was not discrete?

B.H. Yes, but our information is an activity, not a list of instructions. I knew using the word information would cause trouble. It has a number of different meanings. That is why I insisted that we distinguished between the information we were talking about in our book from the information that Shannon introduced. His information has a technical meaning and is not information as we use it in every day language. This is why we can measure it in terms of bits. Shannon information is discrete.

A.G. Could you say more what you mean by information, because some people may say that your concept does not make sense because your information is not discrete, therefore it cannot be measured in bits.

B.H. But that is because we are not using Shannon information. For me the question is tied up with meaning. Are we talking about information for us or are we talking about object information that is independent of any 'knower'? Now, if you look at Shannon information carefully, you find that it is not concerned with meaning. It is more to do with the capacity to carry strings of symbols. The string does not have to mean anything. Thus his information is not what we mean when we say we are giving people information. In fact he pointed out that he was using information in a very technical sense, and made it clear that he was talking about information-capacity rather than information per se. You can provide a measure for information-capacity; you talk about it in terms of bits. This is where discreteness comes in.

Shannon information has nothing to do with information in the sense of 'forming from within' nor does it have any structure to carry meaning. His information has no semantic structure whatsoever and that is why it is discrete. Our information does not have a discrete structure. It is continuous, because we have got the active information carried by the quantum potential, and the quantum potential is continuous.

A.G. Is it wrong to think that information carried by the quantum potential is discrete, but the activity to which it gives rise is continuous?

B.H. How can the quantum potential carry discrete information? I have given some thought to the relationship between the two forms of information and have not come up with any convincing arguments except to say they are different. But I don't see any evidence in quantum theory that this information is discrete. Unless you argue that because the information comes from the wave function and that different wave functions can be orthogonal, these orthogonal wave functions can be the basis of your discreteness.

This way of talking about information is closely related to the density matrix description of quantum phenomena - it has just occurred to me that this is how to tie up the two forms of information. Indeed the quantum potential derived from one of the set of orthogonal wave functions is one bit of information. Another wave function from the set would give another bit and so on. But there is nothing intrinsic in the quantum potential that says that it must be discrete.

WHERE DOES SCHRODINGER'S EQUATION REALLY COME FROM?

A. M. Deakin, 75 Clatterford Rd, Newport, Isle of Wight, PO30 1NZ
E-Mail: anthony.deakin@baedsl.co.uk

ABSTRACT

We begin by summarising a conventional view of Schrödinger's Equation. Questions are asked and this view is criticised. There follows a mathematical experiment wherein an hierarchy of identities is quantised. These identities apply to arbitrary, differentiable functions of the coordinates that are themselves differentiable functions of time; such functions appear in classical dynamics. It turns out that the quantisations are not identities unless we restrict the form of the Hamiltonian operator. The Schrödinger and Dirac forms satisfy whereas arbitrary choices do not. The Dirac form is a better approximation in that it satisfies higher level identities than the Schrödinger form.

1. The Schrödinger Equation

The Schrödinger equation is usually written in the form

(1) $H\phi = ih\dfrac{\partial \phi}{\partial t}$; $h \equiv$ (Plank's constant)$/(2\pi)$

where t is time (treated as a scalar that commutes with all operators), ϕ represents the state and H is the energy operator; H is usually taken to have the same structure as the Hamiltonian of the corresponding classical model (if there is one).

(2) $\phi \equiv \phi(\underline{q},t)$

is taken to be a complex, normed, continuous , differentiable function of the (scalar) coordinates

(3) $\underline{q} \equiv \{q_1, q_2, \ldots q_n\}$

and is defined on an Hilbert space with the $\underline{q}$ as arguments; t is treated as a label. It follows from (1) that

(4) $\phi(\underline{q},t) = \exp(-itH/h)\phi(\underline{q},0)$

The first step, in finding an eigenvalue E of H with eigenfunction ϕ, is to note that

(5) $H\phi = ih\frac{\partial \phi}{\partial t} = E\phi \Rightarrow \phi = \psi(\underline{q})\exp(-itE/h)$

where the eigenfunction $\psi(\underline{q})$ satisfies

(6) $H\psi = E\psi$; also sometimes called Schrödinger's equation

To make further progress we need to specify H as an operator function of other operators with known properties. Typically, when dealing with a particle or system of particles, the operator arguments are the coordinates $\underline{Q}$ and their conjugate momenta $\underline{P}$ with the correspondence (from classical scalars to operators)

(7) $\underline{p} \to \underline{P};\quad \underline{q} \to \underline{Q}$

and the commutation conditions

(8) $P_j P_k = P_k P_j;\quad Q_j Q_k = Q_k Q_j;\quad Q_j P_k - P_k Q_j = ih\delta_{jk} I;\quad j,k = 1,2,\ldots.n$

where I is the unit operator. Recall that if we take the Fourier transforms of the functions $\phi(\underline{q})$, in the Hilbert space in which the operators $\underline{Q}$ are represented by the $\underline{q}I$, we obtain a dual Hilbert space of functions $\vartheta(\underline{p})$. The operators $\underline{P}$, which have the representations $\underline{p}I$ in the dual Hilbert space, can be shown to satisfy (8). Notice that this mathematical fact does not, in itself, justify the interpretation of the $\underline{p} \to \underline{P}$ as momenta.

The original Schrödinger problem (the hydrogen atom) considers the Hamiltonian operator

(9) $H \equiv \frac{1}{2m}\sum_{j=1}^{n} P_j^2 + V(\underline{Q});\quad n = 3;\quad m$ is the scalar mass

with representations

(10) $Q_j \equiv q_j I;\quad P_j \equiv -ih\frac{\partial}{\partial q_j};\quad I$ is the unit operator

which corresponds to a single Newtonian particle in a scalar field $V(\underline{q})$. Here the coordinates are Cartesian; and the representation (10) is suitable only to such coordinates.

2. Criticisms

It is pertinent to ask: where does the form (9) come from and why is it so important? A 'standard' answer is that by integrating Newton's laws, in the context of Euclid's geometry, we arrive at Hamiltonian mechanics; and (9) is simply the operator form of an important Hamiltonian. Whether or not the eigenvalue problem (6) turns out to be useful in QM (quantum mechanics) is a matter for theory and experiment. Another 'standard' answer makes a similar appeal to SR (special relativity) mechanics in which Euclid's geometry and Newton's universal time are replaced by Einstein's space-time. The corresponding eigenvalue equation (6) is Dirac's equation. An interesting point is that, by factorising the SR energy equation, Dirac introduced features that belong, quintessentially, to QM and have no classical analogue.

These 'standard' answers seem to me to be unsatisfactory; we need a deeper insight. Firstly, the efficacy of Newton's laws is profoundly mysterious. Secondly, on the face of it, there is no obvious reason why only some models work both, in the large, as a basis for CM (classical mechanics) and, in the small, as a basis for QM. Thirdly, the structure of space-time, which we all take for granted, is just as mysterious as Newton's laws!

3. The Identities

The following discussion purports to uncover unsuspected links between QM and CM that may shed light on the above questions.

Let $\theta(\underline{q})$ be *any* real, continuous, differentiable function that does not depend explicitly on t. Then, by regarding the coordinates $\underline{q}$ as differentiable functions of time, we have

$$(11) \quad \dot{\theta} = \sum_j \dot{q}_j \frac{\partial \theta}{\partial q_j} + \frac{\partial \theta}{\partial t}; \quad \dot{\theta} \equiv \frac{d\theta}{dt}; \quad \dot{q}_j \equiv \frac{dq_j}{dt}; \quad \frac{\partial \theta}{\partial t} = 0$$

or, in a more compact notation,

$$(11a) \quad \dot{\theta} = \dot{q}^j \theta_{,j}; \quad \theta_{,j} \equiv \frac{\partial \theta}{\partial q^j}; \quad \text{Einstein summation convention in force}$$

where suffices on the coordinates have been raised to implement the summation convention. The identity (11) can be differentiated, with respect to time, to give further 'higher level' identities. Thus

$$(12) \quad \ddot{\theta} = \ddot{q}^j \theta_{,j} + \dot{q}^j \dot{q}^k \theta_{,j,k}$$

and again

$$(13) \quad \dddot{\theta} = \dddot{q}^j \theta_{,j} + 3\ddot{q}^j \dot{q}^k \theta_{,j,k} + \dot{q}^j \dot{q}^k \dot{q}^l \theta_{,j,k,l}$$

Because the $\underline{q}$ are taken to be functions of t we can interpret these differential identities as equations in *classical mechanics.* If (11), (12) and (13) are equations in CM then $\theta(\underline{q})$ must be a variable within a system and, as such, has a physical meaning; but, if we are not prepared to specify either which system or which variable, then $\theta(\underline{q})$ is , for most intents and purposes, arbitrary. That the system is classical is consistent with our assertion that both the $\underline{q}(t)$ and $\theta(\underline{q})$ are differentiable.

4. Quantisation of the First Level Identity (11)

Suppose, first, that we choose $\theta(\underline{q})$ to be arbitrary within its class; and suppose, second, that we attempt to 'quantise' (i.e., provide an operator form for) the classical equation (11) according to the usual QM recipes. The correspondence between classical variables and quantum operators is

(14) (7) given (8); $\theta(\underline{q}) \rightarrow \Theta(\underline{Q})$; $\dfrac{\partial \theta}{\partial t} = 0 \rightarrow O$; O is the null operator

Recall that the expressions for the rate operators follow from the definitions

(15) $a \rightarrow A$; $\dot{a} \rightarrow \dot{A}$

where $\dot{A}$ is defined by

$$(16)\quad \begin{array}{l} \langle t|\dot{A}|t\rangle \equiv \dfrac{d\langle t|A|t\rangle}{dt} \\ \Rightarrow \dot{A} \equiv \dfrac{i}{h}(HA - AH) \end{array} ;\quad |t\rangle = \exp(-itH/h)|0\rangle;\quad \text{see (4)};\quad \frac{\partial A}{\partial t} = O;\quad \forall\ |0\rangle$$

from which it follows that

$$(17)\quad \begin{array}{l} \dot{p}_j \rightarrow \dot{P}_j = \dfrac{i}{h}(HP_j - P_jH) \equiv -\dfrac{\partial H}{\partial Q^j} \equiv -H_{,j};\quad H(\underline{p},\underline{q}) \rightarrow H(\underline{P},\underline{Q}) \\ \dot{q}^j \rightarrow \dot{Q}^j = \dfrac{i}{h}(HQ^j - Q^jH) \equiv \dfrac{\partial H}{\partial P_j} \equiv H^{;j} \end{array}$$

where the PD notation suggests the algorithm required to calculate $\dot{P}_j$ and $\dot{Q}^j$, under the rules (8), given $H(\underline{P},\underline{Q})$; (the order of non-commuting operators must be retained). The results (17), taken together, provide a formal justification for the assumptions: a) that H is the operator that corresponds to the classical Hamiltonian; and b) that the $\underline{P}$ are the momentum operators conjugate to coordinate operators $\underline{Q}$. Again, using (16),

(18) $\dot{\theta} \rightarrow \dot{\Theta} = \frac{i}{h}(H\Theta - \Theta H)$

For the partial derivatives of θ we have

(19) $\frac{\partial \theta}{\partial q_j} \rightarrow \frac{\partial \Theta}{\partial Q_j} = \frac{i}{h}(P_j\Theta - \Theta P_j) \equiv \Theta_{,j}$

where the PD notation suggests the algorithm required to calculate $\Theta_{,j}$, given $\Theta(\underline{Q})$, under the rules (8). Finally, for sums and products, recall that for any two real dynamical variables a and b

(20) $\alpha a + \beta b \rightarrow \alpha A + \beta B; \quad ab \rightarrow (AB + BA)/2; \quad \alpha, \beta$ are real scalar constants

All the above operators are either Hermitian or self adjoint (representing real observables). Notice that we have used the Schrödinger representation in which none of the operators depend on t.

With the above definitions we can express the operator equation that corresponds to (11):

(21) $\dot{\Theta} = \frac{i}{h}(H\Theta - \Theta H) = \frac{1}{2}(H^{;j}\Theta_{,j} + \Theta_{,j}H^{;j})$

5. The Quantisation Restricts H But Allows the Schrödinger Form (9)

Because the function $\theta(\underline{q})$ is arbitrary so is the operator $\Theta(\underline{Q})$. Therefore, because (11) is an identity that holds for arbitrary $\theta(\underline{q})$, (21) should also be an identity that holds for arbitrary $\Theta(\underline{Q})$; but, as it turns out, (21) is an identity *only for certain forms of H*. For example suppose that

(22) $H \equiv P^3; \quad \Theta = \Theta(Q)$

then (21) becomes

(23) $\frac{i}{h}(P^3\Theta - \Theta P^3) = \frac{3}{2}(P^2\Theta' + \Theta' P^2)$

where

(24) $\Theta' \equiv \frac{i}{h}(P\Theta - \Theta P) = \frac{\partial \Theta}{\partial Q}; \quad$ for all $\Theta(Q)$

With the aid of (24) result (23) reduces to

(25) $P^2\Theta' + \Theta' P^2 - 2P\Theta' P = O$

But, also with the aid of (24), we obtain the identity

(26) $P^2\Theta' + \Theta' P^2 - 2P\Theta' P = -h^2\Theta'''$

giving

(27) $\Theta''' = O$

which is not true for arbitrary Θ. So the form (22) is disallowed. If we assume that P and Q commute then (25) becomes an identity. It follows that the RHS of (26) is the error that we make under this (classical) assumption.

On the other hand the choice

(28) $H \equiv P$

yields Θ' for each side of (21); and the choice

(29) $H \equiv P^2$

yields $P\Theta' + \Theta' P$ for each side of (21). We are also allowed

(30) $H^2 \equiv P^2 \Rightarrow H \equiv cP; \quad c^2 \equiv I \Rightarrow c = \pm I$

providing that we take the square root as indicated and provided that c commutes with both P and Q. Each side of (21) then reduces to $c\Theta'$.
By contrast, if we assume

(31) $H^r = P^r \Rightarrow H \equiv cP; \quad c = I^{1/r}; \quad r > 2$

then, the velocity operator

(32) $\dot{Q} = c$; see (17)

is not, in general, Hermitian; so that the velocity is not real and (31) is, therefore, disallowed.

Because (21) is linear in H the success of (29) ensures that the form (9) is allowed. Notice that, because $V(\underline{Q})$ commutes with $\Theta(\underline{Q})$ and disappears from $H^{;j}$, it makes no contribution to (21). Given the form (9) each side of (21) reduces to $\frac{1}{2m}\sum_{j=1}^{3}[P_j\Theta_{,j} + \Theta_{,j}P_j]$.

6. Quantisation of the Second Level Identity (12)- The Schrödinger Form (9) is an Approximation

We could continue to examine the suitability of various other Hamiltonians; but there is a more pressing matter. If quantisation of the first level identity (11) places restrictions on H the same may be true of the quantisations of the second level identity (12). Using the rules of Section 5 we obtain

$$\ddot{\Theta} = \frac{-1}{h^2}[H(H\Theta - \Theta H) - (H\Theta - \Theta H)H]$$

$$(33) \qquad = \frac{i}{2h}[(HH^{;j} - H^{;j}H)\Theta_{,j} + \Theta_{,j}(HH^{;j} - H^{;j}H)]$$
$$+ [(H^{;j}H^{;k} + H^{;k}H^{;j})\Theta_{,j,k} + \Theta_{,j,k}(H^{;j}H^{;k} + H^{;k}H^{;j})]/4$$

The choices (28) and (30) satisfy (33) exactly; but, when we substitute (9), we get

$$(34) \qquad \frac{h^2}{4m^2}\sum_{j,k}\Theta_{,j,j,k,k} = O$$

the LHS of this equation being the imbalance (LHS-RHS) between the two sides of (33). Given that, for (9), $n = 3$ the operator equation (34) corresponds to the scalar PDE

$$(35) \qquad \nabla^2(\nabla^2\theta(\underline{q})) = 0$$

where, we recall, the $\underline{q}$ are the Cartesian coordinates of a single particle. Notice that if the mass m is large and the higher gradients of $\theta(\underline{q})$ are small then the *scalar* imbalance

$$(36) \qquad \frac{h^2}{4m^2}\nabla^2(\nabla^2\theta(\underline{q}))$$

is, physically, very small (i.e., possibly, a very small fraction of $\ddot{\theta}$).

It should be noted that a number of the solutions of (35) (the Poisson equation with a density that is a solution of the Laplace equation) are functions, of importance in CM. For example: the Cartesian coordinates, the radius and its reciprocal, the azimuth (but not the polar angle) and the potential in empty space $V(\underline{q})$ given that $\nabla^2 V = 0$; for these functions, then, (33) is an identity. We might say that (35) selects functions of the coordinates $\theta(\underline{q})$, given (9), for which (33) is satisfied, in addition to (21), by the corresponding operator $\Theta(\underline{Q})$; and, in this sense, (9) represents the dynamical behaviour of these variables more accurately than that of other functions of the coordinates.

7. The Dirac Form (42) is Exact Up to Level Two

The solutions of (35) can approximate many functions; but in no sense are they arbitrary. So, except for a limited class of operators $\Theta(\underline{Q})$, (9) does not satisfy (33). The question arises: can we modify (9) so as still to satisfy (21) while, at the same time, eliminating the error term (see the LHS of (34))? In seeking a suitable modification we must keep in mind that (9) is the quantisation of a classical Hamiltonian

(37) $H(\underline{p},\underline{q})=\frac{1}{2m}\sum_{j=1}^{3}p_j^2+V(\underline{q})+mc^2;$ c is the velocity of light

where the constant mc^2 is usually omitted because it makes no difference to our calculations. But this Hamiltonian is a low momentum approximation to a classical relativistic Hamiltonian given by

(38) $(H(\underline{p},\underline{q})-V(\underline{q}))^2=\sum_{j=1}^{3}c^2p_j^2+m^2c^4$

Taking the square root of (38) we recover (37) thus

(39)
$$H(\underline{p},\underline{q})=mc^2[1+\sum_{j=1}^{3}(\frac{p_j}{mc})^2]^{1/2}+V(\underline{q})$$
$$\approx\frac{1}{2m}\sum_{j=1}^{3}p_j^2+V(\underline{q})+mc^2;\quad mc>>p_j$$

(40) Quantising (38) we get

(41) $(H(\underline{P},\underline{Q})-V(\underline{Q}))^2=\sum_{j=1}^{3}c^2P_j^2+m^2c^4I$

In order to obtain an expression for the operator $H(\underline{P},\underline{Q})$, without approximation, we take the square root of (41) in the manner of Dirac:

(42) $H(\underline{P},\underline{Q})=\sum_{j=1}^{3}c\sigma_jP_j+\sigma_0mc^2+V(\underline{Q})$

(42a) $\sigma_\alpha\sigma_\beta+\sigma_\beta\sigma_\alpha=2\delta_{\alpha\beta}I;\quad \alpha,\beta=0,1,2,3$

where the σ_α are Hermitian operators that commute with both the $\underline{P}$ and the $\underline{Q}$. H, defined by (42) and (42a), satisfies both (21) and (33) exactly. Each side of (21) is equal to

(43) $c\sum_{j=1}^{3}\sigma_j\theta_{,j}$

and each side of (34) is equal to

(44) $2c^3m\sum_{j=1}^{3}\frac{i}{h}\sigma_0\sigma_j\theta_{,j}-c^2\sum_{j\neq k}\frac{i}{h}\sigma_j\sigma_k(P_k\theta_{,j}+\theta_{,j}P_k)+c^2\sum_{j=1}^{3}\theta_{,j,j}$

So, by round about means, the error term has been eliminated.

8. The Dirac Form is Exact At Least Up to Level Three

A further calculation shows that, given the linear form

(45) $H(\underline{P},\underline{Q})\equiv\sigma^jP_j+\sigma+V(\underline{Q})$; Einstein summation convention in force

and provided that the σ^α commute with both the $\underline{P}$ and the $\underline{Q}$, (21) and (33) are *always* identities whatever mutual commutation rules govern the σ^α. It is natural to ask: if we begin with the linear form (45), assuming arbitrary coefficients σ^α, does the third level impose restrictions where the first and second levels impose none?

Unfortunately the calculations involved in the quantisation and subsequent reduction of (13), given (45), are far too voluminous to reproduce here. But it can be shown that the operator equation, thus arrived at, is an identity for arbitrary $\Theta(\underline{Q})$ only if

(46) $(\sigma^j\sigma^k+\sigma^k\sigma^j)\sigma^l=\sigma^j(\sigma^j\sigma^k+\sigma^k\sigma^j)\Rightarrow(\sigma^j)^2\sigma^k=\sigma^k(\sigma^j)^2$; $\forall\, l.j.k$

and

(47) $(\sigma^j\sigma^k+\sigma^k\sigma^j)\sigma=\sigma(\sigma^j\sigma^k+\sigma^k\sigma^j)\Rightarrow(\sigma^j)^2\sigma=\sigma(\sigma^j)^2$; $\forall\, j.k$

These conditions are satisfied by any set of operators σ^j and σ that satisfy

(48) $\sigma^\alpha\sigma^\beta+\sigma^\beta\sigma^\alpha=2\delta^{\alpha\beta}I$; $\alpha,\beta=0,1,2,3$; $\sigma^0\equiv\sigma$; see (42a)

so that the conditions (46) and (47) are less stringent than the conditions (48); perhaps the conditions (48) are imposed at level 4. Note that (46) and (47) were arrived at on the assumption that we are allowed to set

(49) $\sigma=O$

which is like setting

(50) $m=0$; see (42)

9. Conclusions

Two alternative views can be taken of these calculations:

(a) The operator forms of the differential identities themselves fail to be identities for all $\Theta(\underline{Q})$ and all H because the quantisation recipe is at fault.

(b) The insistence that $\theta(\underline{q})$ is a continuous, differential function of the coordinates $\underline{q}$ and, further, that the $\underline{q}$ are continuous, differential functions of the time t identifies $\theta(\underline{q})$, although otherwise arbitrary, as belonging to a classical system. The restrictions placed on H, to ensure that the operator forms remain identities for arbitrary $\Theta(\underline{Q})$, therefore characterise classical systems.

I incline to the second view. Certain well known classical forms for H are here shown to satisfy the restrictions, whereas, arbitrarily chosen forms do not. I claim that the aptness of these classical forms explains the success of both the Schrödinger and the Dirac equations.

It is shown that the Schrödinger form of H can only approximate the higher level identities. The Dirac form is exact at least up to level 3.

It is possible that the hierarchy of operator equations, obtained by successive differentiation of (11) and treated as identities, might determine H in detail; but the technical difficulty of establishing this proposition is obvious. As we ascend the levels the equations become more and more non-linear and the number of terms rises hugely. We would hope, however, that, for certain forms of H, the sequence of conditions so obtained would exhibit closure.

Much work remains in order to decide, finally, between the views (a) and (b). In particular we need an elegant method to extract the conditions at higher levels. This method must avoid my clumsy procedure of differentiation, quantisation, substitution and reduction.

19/5/1999

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

1. The Executive Council is the governing body of the Association. It consists of:

(a) All past presidents of the Association.

(b) Officers (acting president, vice president, treasurer, secretary and coordinator if one is appointed).

(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. Vice-presidential candidates are nominated by the membership and the Executive Council during the first year of the President's term of office. Any nomination must be accompanied by a statement from the nominee that he will serve a full term if elected. If there is more than one nominee, selection will be made by mail ballot to the Membership decided by plurality of votes. The Vice-President is elected to serve concurrently with the President during his last year of office. He will then serve as President for at most five years and cannot run for re-election until three years after his initial term has elapsed. If the President decides to stand down before his five year term has elapsed, he should give the Executive Council one year's notice of his intention, so that a Vice-President may be elected.

4. 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.

5. The Treasurer is the responsible financial officer of the Association for the receipt and disbursement of funds and shall maintain appropriate records of the Association Activities, membership, mailing-lists, etc.

6. The Secretary is responsible for keeping minutes of the Membership and Executive Council meetings, production of a newsletter to keep members of the Association informed of its activities, and such other duties as may be assigned.

7. President, Secretary and Treasurer will not be paid for their services but may, as appropriate, receive funds for travel expenses, secretarial help, etc.

8, The Coordinator, if one is appointed, may be paid an appropriate salary for his services, funds permitting. These services will include the organisation of meetings and the editing of the Proceedings of such meetings for publication, coordination 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.

9. 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.

10. Dues are currently £20.00 per annum.

Executive Council: Dr. John Amson, Dr. Ted Bastin, Mr. Anthony M. Deakin, Dr. Tom Etter, Ms. Arleta Griffor, Prof. Louis Kauffman, Prof. Clive W. Kilmister, Mr. David McGoveran, 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: 0181 594 5064].

Coordinator: At present no coordinator is appointed.

Acting Secretary: Mary Pope, 'Llys Alaw', 10 West End, Penclawdd, Swansea, W. Glam., SA4 3YX, Wales, UK. [Tel.: 01792 850764]

Treasurer: Tony Deakin, 75, Clatterford Road, Carisbrooke, Isle of Wight, PO30 1NZ, UK. [Tel.: 01983 524140]

Newsletter Editor: Arleta Griffor.

Proceedings Editor: Keith Bowden.

Advisory Board. M. Horner (Chairman), Profs. G.F. Chew (Berkeley), C. Isham (Imperial College), M. Redhead (Cambridge), N. Cartwright (LSE), C. W. Kilmister.