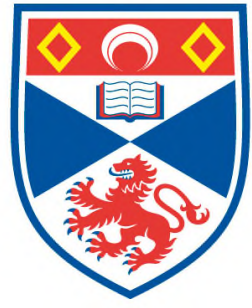
# THE PURSUIT OF LOCALITY IN QUANTUM MECHANICS

**Malcolm Hodkin**

**A Thesis Submitted for the Degree of PhD**
**at the**
**University of St Andrews**



**1990**

# The Pursuit of Locality

# in

# Quantum Mechanics

a thesis by

Malcolm Hodkin

St. Andrews, 1989.

## *A Minor Poet*

The goal is oblivion.
I have arrived early.

Jorge Luis Borges

**Quince Monedas** (Fifteen Coins)
in
**The Gold Of The Tigers**
(trans. Alastair Reid).

I hereby certify that the candidate has fulfilled the conditions of the Resolutions and Regulations appropriate to the degree of Ph. D.

Signed

Date 20 Jun 89

# Abstract

The rampant success of quantum theory is the result of applications of the 'new' quantum mechanics of Schrödinger and Heisenberg (1926–7), the Feynman–Schwinger–Tomonaga Quantum Electro-dynamics (1946–51), the electro-weak theory of Salaam, Weinberg, and Glashow (1967–9), and Quantum Chromodynamics (1973–); in fact, this success of 'the' quantum theory has depended on a continuous stream of brilliant and quite disparate mathematical formulations. In this carefully concealed ferment there lie plenty of unresolved difficulties, simply because in churning out fabulously accurate calculational tools there has been no sensible explanation of all that is going on. It is even argued that such an understanding is nothing to do with physics. A long-standing and famous illustration of this is the paradoxical thought-experiment of Einstein, Podolsky and Rosen (1935).

Fundamental to all quantum theories, and also their paradoxes, is the location of sub-microscopic objects; or, rather, that the specification of such a location is fraught with mathematical inconsistency. This project encompasses a detailed, critical survey of the tangled history of Position within quantum theories. The first step is to show that, contrary to appearances, canonical quantum mechanics has only a vague notion of locality. After analysing a number of previous attempts at a 'relativistic quantum mechanics', two lines of thought are considered in detail. The first is the work of Wan and students, which is shown to be no real improvement on the usual 'nonrelativistic' theory. The second is based on an idea of Dirac's – using backwards–in–time light-cones as the hyersurface in space-time. There remain considerable difficulties in the way of producing a consistent scheme here.

To keep things nicely stirred up, the author then proposes his own approach – an adaptation of Feynman's QED propagators.

This new approach is distinguished from Feynman's since the propagator or Green's function is not obtained by Feynman's rule. The type of equation solved is also different: instead of an initial-value problem, a solution that obeys a time–symmetric causality criterion is found for an inhomogeneous partial differential equation with homogeneous boundary conditions.

To make the consideration of locality more precise, some results of Fourier transform theory are presented in a form that is directly applicable.

Somewhat away form the main thrust of the thesis, there is also an attempt to explain the manner in which quantum effects disappear as the number of particles increases in such things as experimental realisations of the EPR and de Broglie thought experiments.

# Contents

**Chapter**



**Appendicies**

# Chapter 1

## Position, Locality, and thereabouts

*'O what a tangled web we weave,*
*When first we practise to deceive!'*

Marmion by Sir Walter Scott.

## §1 An Introduction

**Making no attempt to be exhaustively precise, Physics may be said to be the study of the natural, inanimate phenomena of fairly simple systems. Generally, these systems are not directly accessible by reason of their distance or size. Physical theory is the result of this study, and encapsulates the best contemporary understanding that has been gained. To be credible, a theory must be at least as precise as the current experimental results, and preferably offer complete precision, so that tests of the theory may be unambiguous.**

One aspect of this, so fundamental and natural as to be almost unnoticed, is the requirement that, for there to be any phenomenon, it must have somewhere and somewhen to happen. It is necessary to convey a sense of the quantitative arrangement of fields, particles and what-have-you that are distinguished by their separation in space-time. This might be called the *concept of Position*. Because position enters most theories at an early stage it has not been often contemplated. Indeed, it is hard to imagine how so basic a concept could be analysed. A simple rule will usually suffice: everything is or happens somewhere in space-time.

To my knowledge, all physical theories comfortably accommodate the concept of position. Where trouble does arise is in a modest development on the postulate that phenomena happen somewhere. Lacking the ingenuity, not to say the budget, to use the whole universe as a tool of scientific enquiry, it is the usual practice for phenomena to be confined within rather more modest, that is finite, bounds of space and time. The prosaic term for this is the *experimental apparatus*. In the normal course of an experiment, steps are taken to eliminate any effect that may be the result of conditions outwith the apparatus. It is only by this means that a scientific investigation makes sense.

A physical theory that successfully mirrors this spatio-temporal limitation is going to be said here to exhibit *locality*. Not all physical theories are this successful: the quantum mechanics of Schrödinger and Heisenberg (among others) is a notorious example of such a failure; it is also on this point that all attempts to formulate a 'relativistic quantum mechanics' come crashing down. This difficulty was clearly enunciated as long ago as 1935, when Albert Einstein, Boris Podolsky and Nathan Rosen insisted that a quantum-mechanical measurement should not affect anything that was space-like separated from the detector.[1] A variety of terms have been borrowed, invented or re-used to describe this facet of what I shall call *the problem of locality;* none are what might be called satisfactory. It will be useful to clear up some of this semantic muddle in preparation for later chapters.

To say, as some do, there is an 'instantaneous spreading of the wave packet' makes as much sense as many discussions of the 'collapse of the wave packet'. That is, not very much. The use of 'instantaneous' suggests its author has not encountered the relativity of simultaneity: those who are better informed are commonly at a loss as to how the measurement postulate of quantum mechanics can be justified for its apparent rubbishing of this relativity.

The collapse of wave-functions as a result of a 'measurement' is uncontroversial if the theory is interpreted entirely in terms of ensembles of 'possible states': the 'collapse' or 'reduction' is then, merely, the selection of a more restricted sub-ensemble. If one hankers for a theory more closely associated with an individual particle there seems no option but to somehow dispense with this discontinuous and non-covariant 'evolution'. A quantum theory exhibiting locality is likely to help here.

Calling some evolution 'causal' or 'acausal' is hardly more apposite. The suggestion made by this is that a particle is, or is not, causing itself. The etymology of this use of 'causal' is from Einstein's Special Principle of Relativity, whereby the envelope of points that may be causally influenced by some event is the light-cone with apex at that event. The deployment of 'acausal' as some sort of negative is contradictory, since the upshot of an 'acausal' propagation is the very possibility of a causal link.

There is a similar perversity in the description of phenomena as 'local' or 'non-local'. In no case is it impossible for the events to be linked by a field, which therefore transfers the influence purely by local action. The distinction that is intended to be made is between the different rates of transfer – specifically, between those occurring at less than, or in excess of, the speed of light. An action–at–a–distance theory, in which the rate of transfer is infinite, could only be called non-local if there is nothing that could be interposed between cause and effect that will be itself affected or will affect this action.

Instead of 'instantaneous' or 'acausal spreading', it would be better to say that there is a

superluminal propagation of the appropriate sort. Not quite as snappy, it is true, but then the intention is to convey a meaning. A typical statement in the approved manner is thus: the support of the wave-function expands at a superluminal rate, contrary to Einstein's Special Principle of Relativity. This will be exact if the Special Principle is taken to be the assumption of light speed as the limit on all forms of propagation.

If a word is needed to carry the negative overtone intended by 'acausal' and the like, then a good choice is *anachoristic*. This is the spatial counterpart to 'anachronistic', meaning the presence of something in the wrong place†. Such terminology must be only temporary, however, for if the universe works in such a way as to manifest phenomena by way of superluminal causes then there is nothing to be in the wrong.

It is interesting that in all the hullabaloo about the experimental tests of Bell's theorem that the results are portrayed as being decisive between 'Quantum Mechanics' and 'Special Relativity'. So far as I know, only Karl Popper has remarked that this is more properly a contest between Einstein's Special Principle and Lorentz's aether interpretation of mechanics. This does still ignore the fact that the quantum predictions arise from a theory invariant with respect to the Galilei group, since there is no satisfactory quantum mechanics invariant under the Poincaré group (the inhomogeneous Lorentz group). It is even curious that no detectable candidates have been offered for the means by which the two arms of the experiment are superluminally connected – the ability to pass unimpeded through metal, concrete and, perhaps, lab technicians may make detection a formidable task. It is in this context that some authors have legitimately used the term 'non-local'; though to do so is to propose only a retrograde step for physical theory, for the rejection of local action is the rejection of an idea of huge explanatory power.

Since I shall be making numerous references to quantum theories with either of the standard symmetry groups, it is convenient to make a further point about nomenclature here. The terms 'nonrelativistic' and 'relativistic' are almost universally accepted qualifiers – used to indicate the symmetry group of a theory. This usage arose, as I understand it, less through deliberation than historical proximity: 'relativistic' came to mean 'invariant under the Poincaré group' simply because the principle of Relativity (in physics) and interest in the Poincaré group‡ came to prominence together. It then became common to describe the older theories as not 'relativistic', whence 'nonrelativistic'. Jean-Marc Lévy-Leblond has pointed out that there is nothing unrelativistic about nonrelativistic theories,[2][3] and that Galilei had even produced a form of relativity principle in connection with what are now known as Galilean transformations or boosts. To avoid repeating the use of these unfortunate epithets, I have adopted a reasonable compromise between clarity and brevity: theories invariant with respect to the Galilei group will be called *Galilei-invariant* or *Galilei-relativistic*; whereas, if the symmetry group is the Poincaré group then the theory will be described as *Poincaré-invariant* or *Poincaré-relativistic*. In a similar way, if a theory is in accordance with Einstein's Special Principle of Relativity, this will be treated as synonymous with Poincaré-relativistic.

If the foregoing sets out the general problem to be tackled, and cuts a little way through the fog of jargon; it remains only to outline the path this Thesis will take.

Chapter 2 : the canonical theory of quantum mechanics (which is Galilei-invariant, of course) is shown to have a well-defined concept of position but only the loosest grasp on locality.

Chapter 3 : the multiple-particle extension of Galilei-invariant quantum mechanics is analysed to see if the use of beams of massive particles in any way influences the standard examples of anachoristic behaviour, viz. de Broglie's paradox and the Einstein–Podolsky–Rosen thought-experiment.

Chapter 4 : surveys the attempts that have been made to produce a Poincaré-relativistic quantum mechanics. This catalogue of honorable failure is nevertheless a guide to the sort of theory that may work.

Chapter 5 : using the complexification of Fourier transform theory, some analyticity conditions are re-derived that make it considerably easier to eliminate theories that lack locality.

Chapter 6 : the recent work of Wan and students on a localised quantum mechanics is explored.

† I am indebted to Jonathan Cole for the re-discovery of this, sadly overlooked, word.

‡ Incidentally, *the* Poincaré group is only a particular instance of the class of groups studied by Henri Poincaré.

Chapter 7 : the recent revival by Derrick of an idea by Dirac for a Poincaré–invariant quantum theory based on backwards–in–time light-cones is thoroughly re-worked and analysed.

Chapter 8 : chastened by the findings of the previous chapters, a new formalism is proposed that is a development of Feynman's version of Quantum Electro-dynamics.

## §2 Bibliography

(References are preceded with the page number on which they first appear.)

2: [1] A. Einstein, B. Podolsky, N. Rosen, *Can a quantum-mechanical description of reality be considered complete?* Phys. Rev. **47** (1935) 777–780.

3: [2] J.–M. Lévy–Leblond, *The pedadgogical rôle and epistemological significance of group theory in quantum mechanics*, Riv. Nuovo Cimento **4** (1974) 99–143.

3: [3] J.–M. Lévy–Leblond, *Galilei Group and Galilean Invariance*, pp 221–299 of *Group Theory and its applications*, vol. 2, edited by Ernest M. Loebl, Academic Press 1971.

# Chapter 2

## Galilei–Invariant Quantum Mechanics and Locality

*An Incompatibility Explored*

*'They sought it with thimbles, they sought it with care;*
*They pursued it with forks and hope;*
*They threatened its life with a railway-share;*
*They charmed it with smiles and soap.'*

The Hunting of The Snark (*an agony, in eight fits*) (1876) by Lewis Carroll.

In this chapter I shall be concerned with the extent to which the evolution of states, obeying the free Schrödinger equation, can be said to be spatially located.

## §1 The Canonical Formalism

The 'new' quantum mechanics, due mainly to Schrödinger and Heisenberg, is readily presented as a series of axioms. This has the advantage of establishing the exact theory with which this Thesis will presently take issue. The axioms here are not von Neumann's elegant and rigorous set, but have been chosen for the explicit manner in which the theory can be presented, and then analysed – there is no essential abandonment of rigour.

**Axiom** *States*

The complete specification of the instantaneous state of a single-particle, quantum, system is a wave-function, $\phi$ , which is an element of the Hilbert space, $L^2\left(\Re^3, d^3x\right)$ .

**Axiom** *Observables*

Physical quantities attributable to a quantum system are represented by some of the self-adjoint operators defined on a dense subset of the space of states.

The *expectation value* of an observable, $\hat{A}$ , for a state, $\phi$ , is defined to be

$$\langle A\rangle = \langle\phi|\hat{A}|\phi\rangle \quad . \tag{1.1}$$

The spectral theorem for self-adjoint operators on a Hilbert space associates a projection–valued measure, $E(\hat{A};\cdot)$ , with each observable ( $\hat{A}$ , here). The last formula can thus be written

$$\langle A\rangle = \int_{a=-\infty}^{\infty} a\, d\langle\phi|E(\hat{A};a)|\phi\rangle$$

whence a probability distribution, $\rho$ , can be deduced for each observable acting on each state:

$$\rho(a) = \langle\phi|E(\hat{A};a)|\phi\rangle \quad .$$

A consequence of the first two axioms is, therefore:

**Corollary** *Max Born's axiom*

The probability density function, representing the likelihood of the presence of the particle in space, is $\rho(x) = |\phi(x)|^2$ .

**Axiom** *Symmetry group: Galilei*

The generators of the symmetry group for the system are the observables for position, $\hat{\mathfrak{x}}$ (boosts); momentum, $\hat{\mathfrak{p}}$ (translations); angular momentum, $\hat{\mathfrak{L}}$ (rotations). The group action is obtained, by way of Stone's theorem, from the unitary operators

$$U(\alpha,\hat{A}) = \exp\left(i\alpha\hat{A}/\hbar\right)$$

for the continuous parameter $\alpha$ , and generator $\hat{A}$ . The following definitions are applied (giving what is called the *coordinate representation*):

$$\begin{aligned}
\left(\hat{\mathfrak{x}}\phi\right)(x) &= x\phi(x) \quad , \\
\hat{\mathfrak{p}}\phi &= -i\hbar\nabla\phi \quad , \\
\hat{\mathfrak{L}} &= \hat{\mathfrak{x}}\wedge\hat{\mathfrak{p}} \quad .
\end{aligned}$$

It is the almost universal practice to omit this last postulate, or to subsume it in the axiom on observables. The definition of the position observable is implied by Born's axiom, but the remainder are generally just a derivation from the symmetry group of the Schrödinger equation. In omitting a symmetry axiom a shorter, and perhaps more elegant, list may be obtained; but, in doing so, an important physical aspect of the theory is obscured. Indeed, in the realm where this theory has its greatest success – the modelling of atoms – the principal interest is symmetry.

**Axiom** *Time Evolution*

An observable, $\hat{H}$ , called the Hamiltonian by analogy with classical mechanics, acts as the generator of evolution in time. The group of time evolutions, $U(t)$ , is obtained from Stone's theorem, as was the case for the symmetry group.

On the domain of $\hat{H}$ , by virtue of the group it generates, Schrödinger's equation is recovered:

$$i\hbar\frac{\partial\phi}{\partial t} = \hat{H}\phi \quad ,$$

and, by the usual manipulation, a continuity equation can then be deduced, which, for this restricted class of states, is one way to express the conservation of probability:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j} = 0 \quad .$$

From Born's postulate and the simple definition of the position observable, the concept of position is so firmly grounded in the standard quantum mechanics that any flaw in the theory must, surely, be caused by other features? Yet despite this uncomplicated appearance of position, this is a theory in which locality can only be extracted by severe constraints. To wit: any state can be confined to a finite spatial volume for a single instant by a 'measurement' of position that ensures the quantum system lies within that volume (thus eliminating from the ensemble any outcome to the contrary). To accomplish such confinement requires:

**Axiom** *reduction of wave-packets*

If a preparatory measurement is performed on a state, $\phi$ , to ensure that the value of some physical quantity, $A$ , lies in an interval $[a,b]$ , then the instantaneous state after the measurement is

$$\frac{E(\hat{A};[a,b])\phi}{\left\|E(\hat{A};[a,b])\phi\right\|}$$

– where $E(\hat{A};[a,b])\phi$ is a projector derived from the spectral function of $\hat{A}$ .

As will be proved in Chapter 5, for a free particle such a spatial confinement lasts no more than the instant of the 'measurement'. The support of wave-functions in the canonical theory is unbounded: it is non-zero almost everywhere and almost everywhen.

There is one result that allows an approximate form of localisation to exist for an extended period of time. It turns out that any freely evolving state with a finite momentum spectrum will eventually lie in a velocity cone defined using the bounds of this momentum spectrum. In mathematics, the 'Asymptotic Localisation Theorem' says:[1]

$$\lim_{t\to\infty}\left\|E\left(\hat{\mathbf{x}};[\mathbf{p}_1 t/m, \mathbf{p}_2 t/m]\right)U(t)\phi\right\| = \left\|E\left(\hat{\mathbf{p}};[\mathbf{p}_1,\mathbf{p}_2]\right)\phi\right\|$$

Therefore, if $E\left(\hat{\mathbf{p}};[\mathbf{p}_1,\mathbf{p}_2]\right)\phi = \phi$ , there is a time, $\tau$ , for any given $\epsilon > 0$ , such that for all future times

$$\left\|E\left(\hat{\mathbf{x}};[\mathbf{p}_1 t/m, \mathbf{p}_2 t/m]\right)U(t)\phi\right\| \geq 1-\epsilon \quad .$$

There is, thus, a form of locality retrieved in the asymptotic limit; or an approximate version to whatever tolerance beyond the time $\tau(\epsilon,\phi)$ . On reflection, the less than satisfactory state of affairs is, perhaps, to be expected of a theory based on the Galilei group, where there is no limit to the speed with which influences may propagate. Only infinite time lapses, leading to infinite separations, will produce the complete disconnection of systems in a Galilei-invariant theory. Indeed, the configuration space inner-product means there is no difficulty in finding an 'observable' that correlates states that are located only a finite distance apart.

## §2 The Configuration Space $S^1$ and Locality

**$S^1$ is the topological space sometimes called the 1-sphere – a closed, one-dimensional space defined by a single parameter, its length, denoted by $\ell$ here. This space is particularly suitable for modelling several important systems, e.g., simple closed universes and 2-arm interferometers. Just as $\Re$ or $\Re^3$ stands in place of an idealised, open, flat universe; so using $S^1$ avoids having to specifically model the means through which particles are reflected around apparatus, and by which path lengths are sometimes controlled. Of course, using $S^1$ there is nothing to stop the particle from going round many times, as there is in practice, between its initial and measured states. The quantum theory of measurement is not seen here at its best: the exclusive consideration of measurements at particular instants of time is hardly a good way of modelling experiments where the position of the detector is the fixed coordinate. Since I am concerned with emulating the approximate localisations that have been found on $\Re$ – whereby, for example, the two components of a de Broglie–type state become distinct – the problem of lapping will not feature here.**

An arbitrary state on $S^1$ has a Fourier series representation:

$$\phi(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x/\ell} \tag{2.1}$$

where

$$c_k = \frac{1}{\ell} \int_{-\ell/2}^{\ell/2} e^{-2\pi i k y/\ell} \phi(y)\, dy \quad . \tag{2.2}$$

By Bessel's inequality, $(\forall N, M)$

$$\|\phi\|^2 \geq \sum_{n=N}^{M} |c_n|^2 \tag{2.3}$$

there are $M = M(\epsilon, \ell)$ and $N = N(\epsilon, \ell)$ such that

$$\left\| \phi - \sum_{n=N}^{M} c_n e^{2\pi i n x/\ell} \right\| < \epsilon \quad . \tag{2.4}$$

Since the Fourier series is also a decomposition of the state in terms of the eigen-vectors of momentum on $S^1$ :

$$\hat{p}\, e^{2\pi i k x/\ell} = \frac{2\pi k \hbar}{\ell}\, e^{2\pi i k x/\ell} \quad ;$$

the truncation of the Fourier series is identical to the confinement of the momentum spectrum to a finite interval. Therefore, if complete accuracy is sacrificed, any wave-function can be roughly reconstructed with a momentum range $[2\pi N\hbar/\ell, 2\pi M\hbar/\ell]$ to within a tolerance of $\epsilon$ (to coin a phrase). Denote this approximation by

$$\phi_N^M(x) = \sum_{n=N}^{M} c_n e^{2\pi i n x/\ell} \quad .$$

The class of states of interest here are those with a support considerably smaller than $\ell$ —or, more precisely, the approximations, $\phi_N^M$ , to these states. A state, $\phi_N^M$ , will be used as the initial state of the system.

Allowing for the evolution of the state in time gives the more general expression:

$$\varphi(x,t) = \sum_{n=-\infty}^{\infty} c_n \exp\left[ \frac{2\pi i n x}{\ell} - \frac{i\hbar t}{2m} \left( \frac{2\pi n}{\ell} \right)^2 \right] \quad .$$

Substituting for $c_n$ using equation 2,

$$= \sum_{n=-\infty}^{\infty} \frac{1}{\ell} \int_{-\ell/2}^{\ell/2} \exp\left[ \frac{2\pi i n (x-y)}{\ell} - \frac{i\hbar t}{2m} \left( \frac{2\pi n}{\ell} \right)^2 \right] \phi(y)\, dy \quad . \tag{2.5}$$

An attempt will now be made to re-cast for $S^1$ the derivation of the Asymptotic Localisation Theorem, as given in the book by Werner Amrein, Josef Jauch and Kalyan Sinha [1]

The space $S^1$ can be regarded as merely a segment of $\Re^1$ if periodic boundary conditions are imposed on a suitable interval. This allows the following estimate to be made:[2]
$(\exists \epsilon > 0)(\exists \ell' : \forall \ell > \ell')$

$$\left\| \left( e^{-itH(\ell)/\hbar} - U(t) \right) \phi \right\| < \frac{\epsilon}{4} \tag{2.6}$$

for all $\phi \in L^2(\Re)$ , uniformly for $t$ in finite intervals of $\Re$ . $H(\ell)$ is the Hamiltonian on $S^1$ embedded in $\Re$ as the interval $[-\ell/2, \ell/2]$ .

While there is no doubt that both equations 4 and 6 can be simultaneously satisfied, there seems no systematic means of producing a $\phi_N^M$ such that

$$\| \phi - \phi_N^M \|_{S^1} < \epsilon \tag{2.7}$$

and $(\forall t \in [0, t''])(\exists \ell' : \ell > \ell')$

$$\left\| \left( e^{-itH(\ell)/\hbar} - U(t) \right) \phi_N^M \right\|_{\Re^1} < \frac{\epsilon}{4} \tag{2.8}$$

– since to find a $\phi_N^M$ consistent with equation 7 depends on the value of $\ell$ ; yet to satisfy equation 8, $\phi_N^M$ must be known in order to set $\ell$ . I will, nevertheless, assume that this juggling act can always be successfully concluded. So the time evolution on $S^1$ can be approximated by that on $\Re^1$ with fixed accuracy for the period $t \in [0, t'']$ —if the range of momentum, as defined on $S^1$, and $\ell$ are taken to be large enough.

The momentum spectrum of $\phi_N^M$ is finite on $S^1$ ; however, it is infinite with respect to $\Re^1$ . To see this, take the Fourier transform to produce the momentum representation of the state, with respect to $\Re^1$ , at $t = 0$ :

$$\begin{aligned} \tilde{\phi} &= \frac{1}{\sqrt{2\pi\hbar}} \int e^{-ipx/\hbar} \phi_N^M(x)\, dx \\ &= \frac{1}{\sqrt{2\pi\hbar}} \sum_{n=N}^{M} c_n \int_{-\ell/2}^{\ell/2} \exp\left( i \left( \frac{2\pi n}{\ell} - \frac{p}{\hbar} \right) x \right) dx \\ &= \frac{1}{\sqrt{2\pi\hbar}} \sum_{n=N}^{M} c_n \frac{2\ell\hbar \sin\left( (2\pi n\hbar - p\ell)/2\hbar \right)}{2\pi n\hbar - p\ell} \end{aligned} \tag{2.9}$$

This is an analytic function – a natural consequence of the finite support of $\phi_N^M$ in $\Re^1$ – as will be discussed further in Chapter 5. Therefore, the momentum spectrum (with respect to $\Re^1$ ) is not finite. Of course, as $\ell \to \infty$ , the momentum representation with respect to $S^1$ (equation 2) becomes progressively more like the momentum representation with respect to $\Re^1$ (equation 9).

By Proposition 3.17 in the book by Amrein, Jauch and Sinha,[1]
$(\forall \phi_N^M \in L^2(\Re))(\forall \epsilon > 0)(\exists t')$ such that $(\forall t > t')$

$$\| (U(t) - C(t)) \phi_N^M \| < \frac{\epsilon}{4} \quad , \tag{2.10}$$

where $C(t)$ is the unitary operator defined by

$$(C(t)\phi_N^M)(x) = \sqrt{\frac{m}{t}} \exp\left( \frac{imx^2}{2\hbar t} \right) \tilde{\phi}\left( \frac{mx}{t} \right) \quad .$$

Whence, given the conditions on equation 8, $(\forall t \in [t', t''])$

$$\begin{aligned} \left\| \left( e^{itH(\ell)/\hbar} - C(t) \right) \phi_N^M \right\| &= \left\| \left( e^{itH(\ell)/\hbar} - U(t) \right) \phi_N^M + (U(t) - C(t))\, \phi_N^M \right\| \\ &\leq \frac{\epsilon}{2} \quad . \end{aligned} \tag{2.11}$$

Now, if $\Lambda$ is a measurable subset of $[-\ell/2, \ell/2]$ – whence also of $S^1$ , by association – then

$$\begin{aligned}\left|\int_\Lambda |(e^{-itH(\ell)/\hbar}\phi_N^M)(x)|^2\,dx - \int_\Lambda |(C(t)\phi_N^M)(x)|^2\,dx\right| \\ = \left|\|E(x;\Lambda)e^{-itH(\ell)/\hbar}\phi_N^M\|^2 - \|E(x;\Lambda)C(t)\phi_N^M\|^2\right| \\ = \left(\|E(x;\Lambda)e^{-itH(\ell)/\hbar}\phi_N^M\| + \|E(x;\Lambda)C(t)\phi_N^M\|\right) \\ \times \left|\|E(x;\Lambda)e^{-itH(\ell)/\hbar}\phi_N^M\| - \|E(x;\Lambda)C(t)\phi_N^M\|\right| \quad . \qquad (2.12)\end{aligned}$$

Since

$$\left\|e^{-itH(\ell)/\hbar}\phi_N^M\right\| = 1 = \|\phi_N^M\| = \|C(t)\phi_N^M\| \quad ,$$

and since $\|E(x;\Lambda)\| \leq 1$ , the triangle inequality implies that the right-hand side of equation 12 is less than

$$2\left\|\left(e^{-itH(\ell)/\hbar} - C(t)\right)\phi_N^M\right\| < \epsilon \quad .$$

If $\Lambda = [\lambda_1, \lambda_2]$ , then the foregoing becomes a relation between momentum and coordinate supports because

$$\int_\Lambda \left|(C(t)\phi_N^M)(x)\right|^2\,dx = \int_{\lambda_1}^{\lambda_2} \left|\tilde{\phi}\left(\frac{mx}{t}\right)\right|^2 \frac{m}{t}dx$$

changing the variable by $p = mx/t$ ,

$$= \int_{p_1}^{p_2} \left|\tilde{\phi}(p)\right|^2\,dp$$

(where $p_j = m\lambda_j/t$ ). The degree of spatial localisation is therefore, eventually, related to the momentum spectrum of the state.

The Asymptotic Localisation result can, therefore, be said to be repeated on $S^1$ to the extent that the momentum spectra of a state with respect to $S^1$ and $\Re^1$ are the same. Essentially, the more $S^1$ looks like $\Re^1$ (i.e., as $\ell \to \infty$ ), the more localisation can be said to occur. In one sense, this is a rather obvious, indeed, inevitable result. The significance lies in the apparent impossibility of improving on the original Asymptotic Localisation Theorem. For it is only as $\ell \to \infty$ that the support of $\tilde{\phi}$ contracts down to a finite interval – the infinity being necessary in this case because $\tilde{\phi}$ is analytic and non-zero for all finite values of $\ell$ .

## §3 Computing The Degree Of Localisation

A more specific method of investigating localisation within $S^1$ is simply to calculate the probability that the particle lies in the velocity cone appearing in the Asymptotic Localisation Theorem. This is made vastly easier by having $S^1$ as the configuration space, since a computer can crunch through the summation required far better than the integrals necessary for the theory on $\Re$ ; the computation can be carried quite far analytically as well. Denoting the velocity cone by the intervals, for variable $t$ , $\{[v_1 t, v_2 t]\}$ ; the probability on this cone, at any instant, is then written:

$$\langle\phi|\chi_{[v_1 t, v_2 t]}|\phi\rangle = \frac{1}{\ell}\int_{v_1 t}^{v_2 t} |\phi|^2\,dx \quad .$$

Now $\phi$ has a decomposition in terms of a Fourier series:

$$\phi(x) = \sum_{n=j}^{j+k} a_n \exp\left(\frac{2\pi i n x}{\ell}\right)$$

at $t = 0$ , where

$$v_1 = \frac{2\pi j\hbar}{m\ell} \quad ,$$
$$v_2 = \frac{2\pi(j+k)\hbar}{m\ell} \quad .$$

So that, for arbitrary $t$ ,

$$\phi(x) = \sum_{n=j}^{j+k} a_n \exp\left(\frac{2\pi i n x}{\ell}\right) \exp\left(-\frac{i\hbar t}{2m}\left(\frac{2\pi n}{\ell}\right)^2\right) \quad .$$

Thus

$$\langle\phi|\chi_{[v_1 t, v_2 t]}|\phi\rangle = \frac{1}{\ell}\int_{v_1 t}^{v_2 t} dx \sum_{n=j}^{j+k}\sum_{n'=j}^{j+k} a_n^* a_{n'} \exp\left(\frac{2\pi i x}{\ell}(n'-n) - \frac{it\hbar}{2m}\left(\frac{2\pi}{\ell}\right)^2\left(n'^2 - n^2\right)\right)$$

The integral can be performed in two parts: one in which $n' \neq n$ and a second for the remainder of the double sum. For the first part the dependence on $x$ disappears. Thence

$$\begin{aligned}
\langle\phi|\chi_{[v_1 t, v_2 t]}|\phi\rangle =& \sum_{n=j}^{j+k} \frac{(v_2 - v_1)t}{\ell}|a_n|^2 + \\
&+ \frac{1}{2\pi i}\sum_{n=j}^{j+k}\sum_{\substack{n'=j \\ n'\neq n}}^{j+k} \frac{a_n^* a_{n'}}{n'-n} \exp\left(-\frac{it\hbar}{2m}\left(\frac{2\pi}{\ell}\right)^2\left(n'^2 - n^2\right)\right)\left[\exp\left(\frac{2\pi i\,(n'-n)}{\ell}x\right)\right]_{x=v_1 t}^{v_2 t} \\
=& \frac{2\pi k\hbar t}{m\ell^2}\sum_{n=j}^{j+k}|a_n|^2 + \\
&+ \frac{1}{2\pi i}\sum_{n=j}^{j+k}\sum_{\substack{n'=j \\ n'\neq n}}^{j+k} \frac{a_n^* a_{n'}}{n'-n} \exp\left(-\frac{i\hbar t}{2m}\left(\frac{2\pi}{\ell}\right)^2\left(n'^2 - n^2\right)\right) \\
&\times \left[\exp\left(\frac{i\hbar t}{m}\left(\frac{2\pi}{\ell}\right)^2 (n'-n)\,(j+k)\right) - \exp\left(\frac{i\hbar t}{m}\left(\frac{2\pi}{\ell}\right)^2 (n'-n)\,j\right)\right]
\end{aligned}$$

Let $s = 2\pi k\hbar t/m\ell^2$ .

$$\begin{aligned}
=& s\sum_{n=j}^{j+k}|a_n|^2 + \\
&+ \frac{1}{2\pi i}\sum_{n=j}^{j+k}\sum_{\substack{n'=j \\ n'\neq n}}^{j+k} \frac{a_n^* a_{n'}}{n'-n}\left[\exp\left(\frac{2\pi i s}{k}(n'-n)\left(-\frac{n'+n}{2} + j + k\right)\right) - \right. \\
&\left. \exp\left(\frac{2\pi i s}{k}(n'-n)\left(-\frac{n'+n}{2} + j\right)\right)\right] \quad .
\end{aligned}$$

Since the state is normalised,

$$\sum_{n'=n=j}^{j+k}|a_n|^2 = 1 \quad ,$$

Now

$$s = \frac{t(v_2 - v_1)}{\ell} \in [0,1] \quad ,$$

which is acceptable as a probability.

For the remaining double summation, if the terms are grouped into pairs invariant under the interchange of $n$ with $n'$, and the $\{a_n\}$ are taken to be real numbers, the contribution to the expectation sum is:

$$\frac{1}{2\pi i}\sum_{n=j}^{j+k}\sum_{n=n'+1}^{j+k}\frac{a_n a_{n'}}{n'-n}\left[i\sin\left(\frac{2\pi s}{k}(n'-n)\left(-\frac{n'+n}{2}+j+k\right)\right)-\right.$$
$$\left.i\sin\left(\frac{2\pi s}{k}(n'-n)\left(-\frac{n'+n}{2}+j\right)\right)\right]$$

(since $\sin A-\sin B=2\cos\frac{A+B}{2}\sin\frac{A-B}{2}$)

$$=\quad\frac{1}{\pi}\sum_{n=j}^{j+k}\sum_{n=n'+1}^{j+k}\frac{a_n a_{n'}}{n'-n}\cos\left(\frac{2\pi s}{k}(n'-n)\left(-\frac{n'+n}{2}+\frac{k}{2}\right)\right)\sin\left(\frac{2\pi s}{k}(n'-n)\frac{k}{2}\right)$$

Thence, introducing the short-hand notation $\langle s\rangle$ for $\langle\phi|\chi_{[v_1 t,v_2 t]}|\phi\rangle$ then

$$\langle s\rangle = s+\frac{1}{\pi}\sum_{n=j}^{j+k}\sum_{n=n'+1}^{j+k}\frac{a_n a_{n'}}{n'-n}\cos\left(\frac{\pi s}{k}(n'-n)(-n'-n+k)\right)\sin\left(\pi s(n'-n)\right)\quad .$$

As $s\to 1$ (i.e., $t(v_2-v_1)\to\ell$) the sine factor in each of the 'cross' ( $n'\neq n$ ) terms tends, eventually, to zero.

The dependence on $\ell$ has been reduced to a subsidiary rôle: as $\ell$ is taken to be larger, so more terms will be needed in the summation to achieve the same level of approximation, i.e., range of momentum.

§§3.1 *An Example Calculation*

The simplest possible example of the sort of initial state alluded to in §2 is the characteristic function on some portion, $[-\alpha\ell,\alpha\ell]$ , of $S^1$ :

$$\phi(x)=\frac{1}{\sqrt{2\alpha}}\chi_{[-\alpha\ell,\alpha\ell]}(x)\quad .$$

whence

$$a_n=\frac{1}{\ell}\int_{S^1}\phi(x)e^{-2n\pi ix/\ell}\,dx\quad ;$$

i.e.,

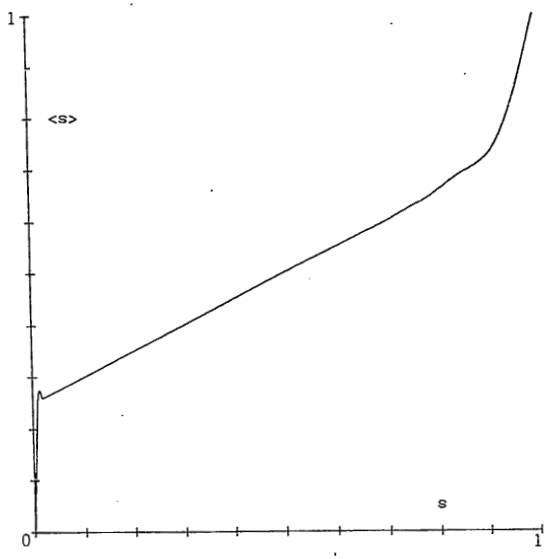
$$a_0=\sqrt{2\alpha}$$

$$a_n=\frac{-1}{\ell\sqrt{2\alpha}}\left[\frac{\ell}{2\pi in}e^{-2n\pi ix/\ell}\right]_{x=-\alpha\ell}^{\alpha\ell}=\frac{\sin(2n\pi\alpha)}{\pi n\sqrt{2\alpha}}=a_{-n}\quad .$$

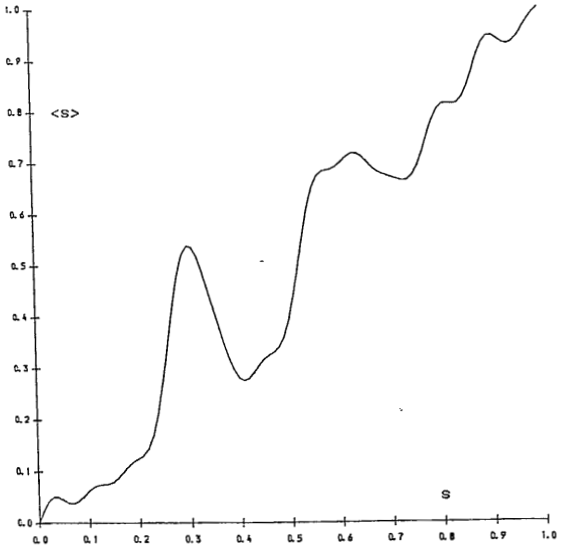
So

$$\phi(x)=\sqrt{2\alpha}+\sum_{n=0}^{N}\frac{2\cos(2n\pi x/\ell)\sin(2n\pi\alpha)}{\pi n\sqrt{2\alpha}}\quad .$$

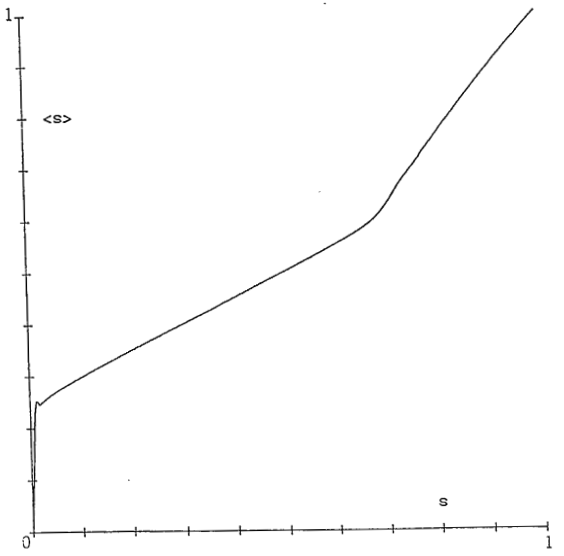
This is illustrated in the next three figures for $\alpha=0.005$ and various values of $N$ .



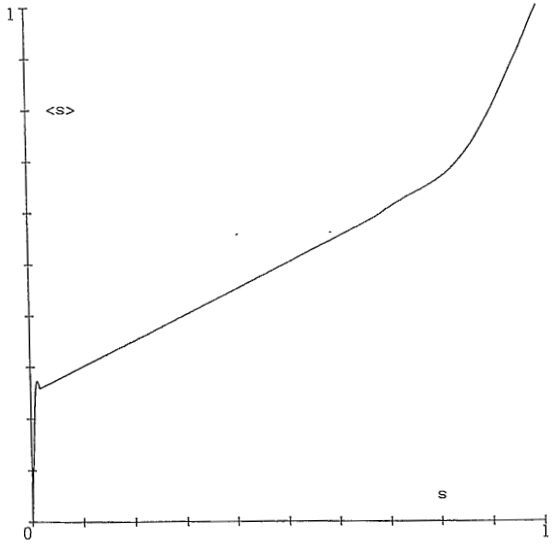

640 terms of : $\langle s \rangle = \left\| E\left(\hat{x}; [v_1 t, v_2 t]\right) \chi_{[-0.005\ell, 0.005\ell]} \right\|$




$a_n = 0$ except: $a_{10} = a_{11} = a_{12} = a_{13} = 1/2$




40 terms of : $\langle s \rangle = \left\| E\left(\hat{x}; [v_1 t, v_2 t]\right) \chi_{[-0.005\mathcal{L}, 0.005\mathcal{L}]} \right\|$




320 terms of : $\langle s \rangle = \left\| E\left(\hat{x}; [v_1 t, v_2 t]\right) \chi_{[-0.005\mathcal{L}, 0.005\mathcal{L}]} \right\|$

## §4 De Broglie's Paradox

This paradox is easily stated in words but lacks any obvious rendition as a formal, quantum-mechanical statement. Roughly speaking, the problem involves a quantum particle placed in a box, which is then divided into two, still sealed, half-boxes. Reason dictates that there is some chance of the particle being in either half, regardless of how far apart they may subsequently be moved. However, on opening one half-box there is a 'collapse of the wave-function' as it becomes certain in which half the particle is. The paradoxical nature of this collapse would not arise for a classical object, where the probability expresses mere likelihood; but for a quantum object the wave-function bears a much closer relation to the nature of the object, implying, perhaps, that a proportion of the particle has instantly evaporated from the empty half-box only to appear in the occupied one.[3]

A straightforward transfer of de Broglie's description into the mathematics of quantum mechanics might be as follows. Take a state, $\phi \in L^2([\mathbf{a},\mathbf{c}])$ , defined on a box, $B$ . The insertion of the divider implies the projection of $\phi$ onto the two half-boxes, $B_1 = [\mathbf{a},\mathbf{b}]$ , $B_2 = [\mathbf{b},\mathbf{c}]$ :

$$\phi = E(\mathbf{x};[\mathbf{a},\mathbf{b}])\phi + E(\mathbf{x};[\mathbf{b},\mathbf{c}])\phi \quad .$$

However, since the two half-boxes represent the same sort of entirely isolated environment as the original box, the process of dividing that box involves more than just the use of the spectral projectors of position. The state of the particle in the two half-boxes must consist of two parts,

$$\phi_L = \hat{P}_L\phi \qquad \text{and} \qquad \phi_R = \hat{P}_R\phi \quad ,$$

where $\hat{P}_L : L^2([\mathbf{a},\mathbf{c}]) \rightarrow L^2([\mathbf{a},\mathbf{b}])$ such that

$$\hat{P}_L\phi = \phi\Big|_{[\mathbf{a},\mathbf{b}]} \quad ;$$

and, similarly, with $\hat{P}_R$ projecting onto $L^2([\mathbf{b},\mathbf{c}])$ .

$\phi_L$ and $\phi_R$ lie in different Hilbert spaces, so the state of the particle is now a *mixture*. But a mixture of states is the same as a classical 'mixture': the particle can now be said to be in either $B_1$ or $B_2$ . The collapse of the wave-function is, in this case, not at all mysterious; for the same reason that the use of a classical object (for example, a pea) in the same situation can only be definitely located in $B_1$ or $B_2$ by opening one of these half-boxes.

The straightforward consideration of de Broglie's 'paradox' is, therefore, only paradoxical if it is asserted that the state measures the actual presence of the particle. Since there is no inconsistency in interpreting quantum mechanical mixtures in the same manner as classical 'mixtures' (a set of mutually exclusive events, each with a probability), then it is reasonable to use a less strict interpretation of quantum mechanical states.

A less exact rendition of the situation proposed by de Broglie avoids splitting the state into a mixture by the use of an instantaneous and discontinuous process.[4] To do this, the boxes are discarded; leaving a particle in the configuration space $L^2(\Re^3)$ .

To simulate the initial confinement in a single box, the particle can be given an initial state, $\psi$ , with compact support, i.e., $\operatorname{supp}\psi(\mathbf{x},0) = [\mathbf{a},\mathbf{c}]$ . Since $\psi(\mathbf{x},0)$ is bounded spatially, its momentum spectrum is the restriction of an analytic function (this is a topic covered more thoroughly in Chapter 5, below). In other words, the momentum spectrum of $\psi$ is the whole of $\Re^3$ , with the possible exception of a set of measure zero. Such a state does not evolve in any way that might simulate the splitting of the original box in half. If $\psi$ is only approximately confined to $[\mathbf{a},\mathbf{c}]$ , i.e.,

$$\left\| E(\hat{\mathbf{x}};[\mathbf{a},\mathbf{c}])\psi \right\| \geq 1 - \epsilon \quad ,$$

but has $\Re^3$ as its spatial support; then it is possible for the momentum spectrum to consist of two bounded but disjoint intervals. Thus

$$\psi = \psi' + \psi''$$

where

$$E(\hat{\mathbf{p}};[\mathbf{p_1},\mathbf{p_2}])\psi' = \psi' \quad ,$$
$$E(\hat{\mathbf{p}};[\mathbf{p_3},\mathbf{p_4}])\psi'' = \psi'' \quad ;$$

and where the components of $\mathbf{p_2} - \mathbf{p_1}$ , $\mathbf{p_3} - \mathbf{p_2}$ , $\mathbf{p_4} - \mathbf{p_3}$ are all positive definite.

The Asymptotic Localisation Theorem can now be applied to deduce that, beyond some time, $\tau$, the two components of the superposition will be approximately confined within well-separated intervals:

$$\|E(\hat{\mathbf{x}};[\mathbf{p_1}t/m,\mathbf{p_2}t/m])U(t)\psi'\| \geq 1-\epsilon$$
$$\|E(\hat{\mathbf{x}};[\mathbf{p_3}t/m,\mathbf{p_4}t/m])U(t)\psi''\| \geq 1-\epsilon \quad .$$

The process of dividing the original box and separating the two half-boxes may now be assumed to occur over the period $0 < t < \tau$ .

It might be not unreasonable to suppose that an observation on a state should not be greatly influenced by the small 'tails' outwith the region of approximate confinement. The observables allowed by this supposition are called *local observables* (this will be dealt with in more detail in Chapter 6). If this is the only set of observables then, not only are $\psi'$ and $\psi''$ orthogonal, but for any local observable, $\hat{A}$ , measuring some aspect of $\psi'$ will give

$$\langle\psi''|\hat{A}|\psi''\rangle < \epsilon$$

and

$$|\langle\psi'|\hat{A}|\psi''\rangle| < \epsilon$$

(which I shall call the *overlap* between $\psi'$ and $\psi''$, with respect to $\hat{A}$ ) and *vice versa*.

With respect to the class of local observables, it therefore follows, the two states, $\psi'$ and $\psi''$, belong to nearly disjoint sub-spaces, i.e., they form an approximate mixture for times later than $\tau$. De Broglie's paradox has been formally presented, and is now resolved by the argument used in the 'straightforward approach', above. The Problem of Locality has also been illustrated.

## §5 A Starting Point

For both $S^1$ and $\Re^1$ it is crucial that localisation may occur, rather than that states may be, in any substantial way, localised. The locality of states is only absolute in the limit of infinite times. The fact that the Asymptotic Localisation Theorem has to use the strong, rather than the uniform, topology on the Hilbert space means that there will be states that take longer than any given time to localise within the velocity cone to the stated tolerance. This contrasts sharply with the fact that, experimentally, there is no great difficulty in confining quantum systems in finite volumes for long periods of time.

It is, I think, undeniable that experiments and phenomena on the sub-microscopic scale are more localised than the Schrödinger–Heisenberg quantum mechanics implies. It is not acceptable to assert, in what amounts to an *ad hoc* alteration of the canonical theory, that 'infinity means macroscopic': that only infinite distances and times take one out of the quantum realm and into the domain of classical physics, where we poor limited creatures can distinguish any details. However, if there were some way to distinguish states that localise to a velocity cone at different rates – a classification theorem specifying a range of closed sub-spaces, for example – then it might be possible to find a physical justification for excluding those states that localise more slowly. In my limited experience, this mathematical result does not exist; the problem of trying to justify approximately confined states does not, therefore, arise. It appears that the concept of locality does not fit into the canonical quantum mechanics.

The problem of locality has a further implication that will be pursued in subsequent chapters. In classical point mechanics, the momentum of a particle and its mass determine exactly the trajectory of that particle. In quantum mechanics, only when Asymptotic Localisation begins to set in does the space-time evolution of a state begin to respect the momentum spectrum of the particle. For finite passages of time, it follows, the momentum spectrum is only rather loosely connected to the dynamics of the particle. Indeed, it would appear to be inappropriate to say the particle was moving in any direction at all until there is some form of approximate localisation to its state. This is not to suggest, necessarily, that the momentum of a quantum particle is not tied strongly to the dynamical behaviour of that particle; but the canonical theory does not show this well.

To me, the obvious direction to investigate is a quantum theory founded on the Poincaré group. In the limitation of propagation to below the speed of light, which is associated with this group,

lies the best hope for any vestige of locality to be discovered. Too late, I learn that such naïve, apparently sensible, thoughts tend to get their thinkers metaphorically chained to eagle-infested rocks.

## §6 Bibliography

(References are preceded with the page number on which they first appear.)

3: [1] Derivations of the Asymptotic Localisation Theorem:

J. D. Dollard, *Scattering Into Cones*, Comm. Math. Phys. **12** (1969) 193–203.

W. O. Amrein, J. M. Jauch, K. B. Sinha, *Scattering Theory in Quantum Mechanics*, W. A. Benjamin 1977, pp 120–125.

M. Reed, B. Simon, *Methods of Modern Mathematical Physics*, vol. II, *Fourier Analysis, Self-Adjointness*, Academic Press 1975; Theorem IX.31.

R. S. Strichartz, *Asymptotic Behaviour of Waves*, J. Func. Anal. **40** (1981) 341–357.

V. Enss, *Asymptotic observables on scattering states*, Comm. Math. Phys. **89** (1983) 245–268.

4: [2] O. Bratteli, D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics, I*, Springer–Verlag 1979; Example 3.1.29, p 187.

11: [3] L. de Broglie, *L'interprétation de la mécanique ondulatoire*, J. de Phys. et Radium **20** (1959) 963–979. See the appendix to this chapter for a translation of §2.

11: [4] K. K. Wan, R. G. D. McLean, *An algebraic approach to quantum mechanics and the EPR paradox*, Phys. Lett. **102a** (1984) 163–166.

### Appendix: L. de Broglie, *L'interprétation de la mécanique ondulatoire*, Le Journal de Physique et le Radium **20** (1959) 963–979. *Translated by Gavin Park*

*§2 The Resolution of Difficulties in the Contemporary Interpretation of Wave Mechanics*

The strongest objections raised against the contemporary interpretation of quantum mechanics are concerned with the non-localisation of the particle in this interpretation. It states, effectively, that if our knowledge of a particle is represented by the wave packet, $\psi$ , the particle is present at all points of this wave packet with probability $|\psi|^2$ . This presence could be qualified by 'potential', and only at the moment when the particle is determined to be at a point in the wave packet by observation does this potentiality become concrete, philosophically speaking. Such a view runs into difficulties that were forcefully indicated by Messrs Einstein and Schrödinger. I have recently reworked my analysis of this kind of difficulty in a book about von Neumann's theory of measure.

These objections may be presented in many different ways. I shall restrict myself to the development of one – a little sketchy – but which shows the nature of the paradoxes that arise. Consider a particle shut in a box, $B$ , which has impenetrable sides. Its wave packet, $\psi$ , is spread throughout the box, so the particle is 'potentially' anywhere in the box with a local probability $|\psi|^2$ . Suppose that by some process, e.g., sliding a divider across the box, it is divided into two isolated parts, $B_1$ , $B_2$ , and then the two boxes, $B_1$ and $B_2$ , are taken to far-distant locations – Paris and Tokyo. The particle is still potentially present in the whole of $B_1$ and $B_2$ , and the wave-function, $\psi_1$ , comprises both $\psi_1$ , localised in box $B_1$ , and $\psi_2$ , localised in $B_2$ . $\psi$ is therefore of the form

$$\psi = c_1\psi_1 + c_2\psi_2$$

where $c_1$ and $c_2$ are constants, generally complex, such that $|c_1|^2 + |c_2|^2 = 1$ .

The laws of probability in wave mechanics say that if an experiment were performed in Paris on box $B_1$ that would determine the existence of the particle in the box, the probability of the result being positive is $|c_1|^2$ and being negative is $|c_2|^2$ . According to the usual interpretation, this has the following meaning :– the particle, 'potentially' present in the whole of the box before the localisation experiment, will suddenly be localised in box $B_1$ , if the result is positive, and in box $B_2$ , if the result

is negative. I do not find such a point of view feasible. The following is the only explanation that seems reasonable to me: the particle was, *before* the localisation experiment, in one of the two boxes, but we do not know which, and the probabilities propounded by the usual view of wave mechanics expresses this ignorance; if we determine that it is in $B_1$ this is because it was already there, and if we could not find it then it was in $B_2$ . Thus everything is clear because we have returned to the classical interpretation of probability the intervention of which results from our ignorance. But as soon as you accept this point of view, it appears that the description of the particle by the wave function, $\psi$ , although leading to a perfectly exact representation of probabilities, does not give a *complete* description of the physical reality, since the particle must have a localisation before the experimental determination and the wave, $\psi$ , is silent about this.

The example developed is a little sketchy, but many more of them may be found: the same problems are always found in different guises. Nothing is gained by using von Neumann's formalism of statistical matrices, this formalism adds nothing to the principles of the probabilistic interpretation of wave mechanics, and if it is applied to the above example, it is clear that the problem is inherent: I have considered the von Neumann theory elsewhere and shown that it falls into paradoxes related to non-localisation, just like the primitive formalism. As for quantum field theory, this contains more than the primitive formalism of wave mechanics since it can represent the constant interaction of charged particles with an electro-magnetic field as well as the creation and annihilation of particles; it cannot alleviate the difficulties in question: in the foregoing example, the intervention of phenomena which may be represented by quantum field theory does not explain how an experiment in Paris results in the localisation of a particle , either in Paris or in Tokyo, which was previously not localised in either place.

The fact that everything in the physical world is localised at each instant in time in the specified frame of reference is fundamental to our experience: the introduction of the concepts of the theory of Relativity and Einsteinian space-time does not change this conclusion. Abandoning localisation does not allow any imaginable picture of the physical world, and that is a consequence which is so grave that there is cause to try everything to avoid it.

Besides, there are other difficulties in the current interpretation of wave mechanics; notably concerning the characteristic nature of the wave packet, $\psi$ , as it is usually considered. It is effectively impossible to consider this wave packet, $\psi$ , as having the concrete nature of physical reality which is attributed to vibrations in classical physics. The possibility of normalising the wave packet, $\psi$ , by arbitrarily choosing its amplitude, the need to change this wave when new data of the state of the particle are obtained, lead naturally to the use of a simple representation of probability lacking an objective character. But this totally subjective view of the wave packet, $\psi$ , involves huge problems: these are interference, which dictates the possible position of the particle, and it is rather difficult when, on reflection, for example, on the phenomenon of electron diffraction, not to accept that the wave propagating through space has a physical reality. Furthermore, the quantised states of atomic systems, to which the characteristic of physical reality must be attributed, are determined by the fact that the associated waves are standing waves with frequencies determined, as for standing waves in classical mechanics, by a calculation of *actual* values. All this does not suggest that waves, $\psi$ , are a purely subjective representation of probabilities: *there must be something objective*. We should be wary of being too insistent on this point in a treatise on quantum mechanics or wave mechanics: authors seem to swing endlessly between the idea that the wave packet, $\psi$ , is a simple representation of probability, and the idea that it has a physical reality. As a teacher of wave mechanics for more than thirty years, I know that even I have constantly performed this kind of swing.

I have insisted on the difficulties presented by the non-localisation of the particle and the subjective character of the wave packet, $\psi$ , in the contemporary interpretation, but thus far I have not spoken about the indeterminism introduced at the same time through quantum physics, which is introduced almost by necessity since asserting determinism is to establish a chain of relations in the form of space and time in such a way that abandoning localisation leads to the abandonment of determinism. But the objections, following Einstein and Schrödinger, that I now see in the purely probabilistic account of wave mechanics are more like non-localisation than the absence of determinism: localised particles can be imagined which exhibit quite indeterminate motion, and the problems indicated above in the example of the boxes, $B_1$ and $B_2$ , would disappear. However, establishing determinism, or causality (the two terms are quite difficult to differentiate), is true to the traditional path of scientific thought. The Double Solution Theory, which I will discuss below, re-establishes determinism at the same time as localisation; but it must, as we shall see, also introduce an element of uncertainty which could be connected to a hidden determinism. But beyond all philosophical discussion of determinism or causality, the essential point for me is still the re-establishment of localisation and objectivity.

# Chapter 3

## Chronological Disordering

*an outline*

*'A freckled and frivolous cake there was*
*That sailed on a pointless sea,*
*Or any lugubrious lake there was*
*In a manner emphatic and free.*
*How jointlessly, and how jointlessly*
*The frivolous cake sailed by*
*On the waves of the ocean that pointlessly*
*Threw fish at the lilac sky.'*

**The Frivolous Cake** by Mervyn Peake.

## §1 Introduction

The Einstein-Podolsky-Rosen and de Broglie thought experiments involve only two and one particles, respectively. If, in an experimental realisation of either of these thought experiments, only a single particle, or pair of particles, is used then there is a direct correspondence between theory and results. On the other hand, it is not uncommon for such experiments to involve beams of particles. To reconcile such experiments and the usual theory it is necessary to assume that the experimental results are not influenced by the several sets of states present. Whether this *independent-particle hypothesis* is justified is a matter for each experimenter to determine: if a configuration can be found so that the particle beams can be treated as a series of independent experiments then there is no problem; in the absence of such a configuration a more sophisticated theory is required. As yet there are no very substantive results for massive particles involved in EPR- or de Broglie-type experiments, but, in anticipation of such work, a sketch of this 'more sophisticated' theory will be given.

To represent a beam of $N$ particles, each particle is assumed to be in the same state except for a time translation: if the superscript denotes the order in which the particles enter the experiment and the subscript denotes the number of particles represented by the state, then

$$\begin{aligned}\phi_1^{(i+1)}(\mathbf{x},t) &= U(-\tau)\phi_1^{(i)}(\mathbf{x},t)\\ &= \phi_1^{(i)}(\mathbf{x},t-\tau)\end{aligned} \tag{1.1}$$

where $U(\tau)$ is the unitary (free) time evolution that takes a state at time $t$ to one at time $t+\tau$. In order to employ arguments based on asymptotic localisation, it is further assumed that the momentum spectrum of the states is bounded:

$$\phi_1^{(i)} = \hat{E}(\hat{\mathbf{p}};[\mathbf{p}_1,\mathbf{p}_2])\phi_1^{(i)} \qquad (\forall i \in \{1,\ldots,N\}) \tag{1.2}$$

All the beams discussed in this chapter will conform to these conditions, or simple variations of them.

The notion of *Chronological Disordering* can then be stated as follows:

> If at some time, $t_0$, the particles in the beam are 'well-separated' then as the states of the particles undergo asymptotic localisation this separation will be eroded. Account will then have to be taken of the indistinguishability of the particles. Therefore, the simple, independent-particle hypothesis, with its 1–particle expectation values computed from 1–particle states, is no longer applicable.

Very roughly, chronological disordering is the effect on expectation calculations arising from the fact that the order of emission of a beam may not be the order in which its constituent particles are later detected.

If chronological disordering has an adverse effect on a single beam, its effect on two beams, where the particles are correlated in pairs between the beams, is liable to be as drastic.

## §2 $N$–Particle Beams : $M$–Particle Observables

### §§2.1 *States*

To represent the quantum theoretic state of a beam of $N$ particles (whether distinguishable or not) one takes the $N$ Hilbert spaces of single particle states, $\mathcal{H}_1$, and forms the tensor product space, also a Hilbert space:

$$\mathcal{H}_N = \mathcal{H}_1 \otimes \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_1 \qquad (N \text{ terms}).$$

Now, an arbitrary state in $\mathcal{H}_N$ may not be 'factorizable', i.e., it may not be of the form

$$\Phi = \phi_1' \otimes \phi_1'' \otimes \ldots \otimes \phi_1^{(N)}$$

but may be a linear combination of such factorized states. This is immediately the case if some or all of the particles are indistinguishable; if some or all obey fermi or bose statistics. For convenience, the 1–particle states in each component term of an $N$–particle state will be numbered from left to right as the '$1^{st}$' to '$N^{th}$ factors'.

The 1–particle factors in any term of an $N$–particle state represent states for each particle which can co-exist. Thus, for example, the sum of the 1–particle spin expectation values in the $z$-direction add up to an allowed value for the experiment modelled. So when the terms of an $N$–particle state are written down they already reflect the various multi-particle conservation laws and constraints on the system based on 1–particle observables.

The construction of an $N$–particle state to model a specific system is something that has not really been systematically expounded. The space of all states is prescribed above; and there are rules about symmetrising or anti-symmetrising with respect to coordinate interchanges. Beyond these two, certainly potent, rules, writing down the state of a system is largely a matter of guesswork. This has already been illustrated in Chapter 2, where turning de Broglie's verbal paradox into a formal statement proved only partially successful.

§§2.2 *Observables*

In a similar fashion to the construction of $N$–particle states, there are $N$–particle observables:

$$\hat{A}_N = \hat{A}_1' \otimes \hat{A}_1'' \otimes \ldots \otimes \hat{A}_1^{(N)}$$

though if some of the $\{ \hat{A}_1^{(i)} \}$ are identity operators such an observable is concerned with fewer than $N$ of the particles. Call any $\hat{A}_N$ with $M$ factors that are not $\hat{I}_1$ an '$M$–particle observable' (on an $N$–particle system). Again, $\hat{A}_N$ is not necessarily factorizable – it may consist of a linear combination of factorizable observables –, and will be symmetrized with respect to the indistinguishable particles of the system – though never anti-symmetrized, which would turn bosons into fermions and *vice versa*.

There is thus a prescription for writing down a general observable on an $N$–particle system given $N$ 1–particle observables. But is this definition complete or sufficient? The constituent 1–particle observables may be fine, as such, but whether the composite of these is valid, i.e., self-adjoint, for an $N$–particle system is a non-trivial question.

As an example on which to fix, take a 2–particle state, $\Psi_2 = \phi_1 \otimes \psi_1 + \psi_1 \otimes \xi_1$ , and the 2–particle observable $\hat{A}_2 = \hat{\mathbf{p}} \otimes \hat{\mathbf{x}}$ :

$$\begin{aligned}\langle \Psi_2 | \hat{A}_2 | \Psi_2 \rangle = {} & \langle \phi_1 | \hat{\mathbf{p}} | \phi_1 \rangle \langle \psi_1 | \hat{\mathbf{x}} | \psi_1 \rangle + \langle \psi_1 | \hat{\mathbf{p}} | \psi_1 \rangle \langle \xi_1 | \hat{\mathbf{x}} | \xi_1 \rangle + \langle \phi_1 | \hat{\mathbf{p}} | \psi_1 \rangle \langle \psi_1 | \hat{\mathbf{x}} | \xi_1 \rangle \\ & + \langle \psi_1 | \hat{\mathbf{p}} | \phi_1 \rangle \langle \xi_1 | \hat{\mathbf{x}} | \psi_1 \rangle\end{aligned}$$

Thus, if an $N$–particle observable is made up from valid (i.e., self-adjoint) 1–particle observables, then there seems no reason to doubt that the composite observable is well-defined. That this does not necessarily give an exhaustive characterisation of $N$–particle observables – at least there is no proof that it does – is not going to be pursued further here. All $N$–particle observables examined in the sequel are constructed of self-adjoint 1–particle operators.

## §3 A Formulation of Chronological Disordering

In the remarks that follow a single beam of $N$ particles will be discussed: $\{ \phi_1', \phi_1'', \ldots, \phi_1^{(N)} \}$ . All that is necessary to obtain the corresponding formulation for two beams is a set of single-particle states, $\{ \chi_1', \chi_1'', \ldots, \chi_1^{(N)} \}$ , that obey the same constraints but for a momentum spectrum $[-\mathbf{p}_2, -\mathbf{p}_1]$ rather than $[\mathbf{p}_1, \mathbf{p}_2]$ .

The beam $\{ \phi_1^{(i)} \}$ is constructed of $N$ 1–particle states that obey (for some $\tau = \tau(\epsilon_1, \epsilon_2)$ ), following the argument of Chapter 2, §4:

$$\| \phi_1^{(i)} \| = 1 \tag{3.1}$$

$$\phi_1^{(i+1)} = U(-\tau)\phi_1^{(i)} \tag{3.2}$$

$$\phi_1^{(i)} = \hat{E}(\hat{\mathbf{p}};[\mathbf{p}_1,\mathbf{p}_2])\phi_1^{(i)} \qquad (\forall i \in \{1,\ldots,N\}) \quad ; \tag{3.3}$$

At $t = l\tau$ ,

$$\|\hat{E}(\hat{\mathbf{x}};\Lambda)\phi_1^{(l)}\| \geq 1-\epsilon_1 \quad ; \tag{3.4}$$

$$(\forall t \geq (l+1)\tau) \qquad \left\|\hat{E}\left(\hat{\mathbf{x}};\left[\frac{t\mathbf{p}_1}{m},\frac{t\mathbf{p}_2}{m}\right]\right)U(t)\phi_1^{(l)}\right\| \geq 1-\epsilon_2 \quad . \tag{3.5}$$

In order to, at least initially, satisfy an independent-particle hypothesis, an additional constraint is imposed on $\tau$ (possibly making it larger):

$$\Lambda \cap \left[\frac{\mathbf{p}_1\tau}{m},\frac{\mathbf{p}_2\tau}{m}\right] = \emptyset \quad . \tag{3.6}$$

Denote by $\Lambda'$ the complement of

$$\Lambda \cup \left[\frac{\mathbf{p}_1\tau}{m},\frac{\mathbf{p}_2\tau}{m}\right]$$

in the configuration space. Whence, at $t = \tau$ , when $\phi_1''$ takes its initial value and $\phi_1'$ is $\tau$ seconds evolved from its initial value,

$$\begin{aligned}|\langle\phi_1''|\phi_1'\rangle| &= \left|\left\langle \phi_1'' \,\middle|\, E\left(\hat{\mathbf{x}};\Lambda\cup\left[\frac{\mathbf{p}_1\tau}{m},\frac{\mathbf{p}_2\tau}{m}\right]\cup\Lambda'\right)\phi_1'\right\rangle\right| \\ &\leq |\langle\phi_1''|E(\hat{\mathbf{x}};\Lambda)\,\phi_1'\rangle| + \\ &\qquad \left|\left\langle\phi_1''|E\left(\hat{\mathbf{x}};\left[\frac{\mathbf{p}_1\tau}{m},\frac{\mathbf{p}_2\tau}{m}\right]\right)\phi_1'\right\rangle\right| + |\langle\phi_1''|E(\hat{\mathbf{x}};\Lambda')\,\phi_1'\rangle| \\ &\leq \epsilon_2+\epsilon_1+\epsilon_1\epsilon_2 \quad .\end{aligned}$$

(By equations 4, 5, the Cauchy-Schwartz, and triangle inequalities.)

And take

$$\epsilon = \epsilon_1+\epsilon_2+\epsilon_1\epsilon_2 \quad .$$

The independent-particle hypothesis is thus true at the $\epsilon$-level (to a tolerance of $\epsilon$) for this 2-particle beam at time $\tau$ ; it is interesting to note the effort needed to satisfy this hypothesis even approximately. To extend the (initial) validity of the hypothesis to an $N$-particle beam, it is necessary that the approximate localisations of the $N-1$ particles, which have developed from their common 'initial value', be disjoint; by taking $\mathbf{p}_2 > \mathbf{p}_1 > 0$ the condition on $\Lambda$ (equation 6) will automatically apply to all $N-1$ approximate localisations of the particles $\phi_1',\ldots,\phi_1^{(N-1)}$ . To obtain an approximate separation, the boundaries of the respective approximate localisations must be well-ordered:

$$\frac{\mathbf{p}_1\tau}{m} < \frac{\mathbf{p}_2\tau}{m} < \frac{2\mathbf{p}_1\tau}{m} < \frac{2\mathbf{p}_2\tau}{m} < \ldots < \frac{(N-1)\mathbf{p}_1\tau}{m} < \frac{(N-1)\mathbf{p}_2\tau}{m} \tag{3.7}$$

Fortunately, this follows immediately if only the penultimate inequality is true:

$$\frac{(N-2)\mathbf{p}_2\tau}{m} < \frac{(N-1)\mathbf{p}_1\tau}{m}$$

or

$$\mathbf{p}_2 < \frac{(N-1)}{(N-2)}\mathbf{p}_1 \tag{3.8}$$

since, if $n \in \{1,\ldots,N-2\}$

$$2 \geq \frac{n+1}{n} \geq \frac{N-1}{N-2} > 1 \quad .$$

Note that this condition fails in the limit $(N \to \infty)$, as any independent-particle hypothesis probably ought.

The constraint on the momentum spectrum of the $\phi_1^{(i)}$ can be summarised as

$$0 < \mathbf{p}_1 < \mathbf{p}_2 < \left(\frac{N-1}{N-2}\right)\mathbf{p}_1 \quad . \tag{3.9}$$

To show the effect of the passage of time on this approximate separation, time will be measured as a multiple, $t\tau$ , of $\tau$ , and the set of boundaries being compared will be extended to include the $N^{th}$ particle ($\phi_1^{(N)}$):

$$t\tau\mathbf{p}_1 < t\tau\mathbf{p}_2 : (t+1)\tau\mathbf{p}_1 < (t+1)\tau\mathbf{p}_2 : \ldots : (t+N-1)\tau\mathbf{p}_1 < (t+N-1)\tau\mathbf{p}_2 \quad .$$

Now, just as $\mathbf{p}_2$ can be bracketed by $\mathbf{p}_1$ and a multiple of $\mathbf{p}_1$ , so there is a positive integer, $s_1$ , such that

$$\left(\frac{N-2}{N-1}\right)\mathbf{p}_2 < \mathbf{p}_1 < \left(\frac{N-2+s_1}{N-1+s_1}\right)\mathbf{p}_2 \tag{3.10}$$

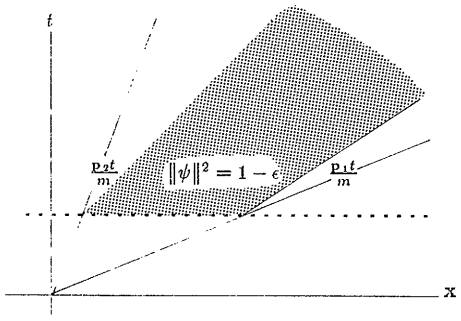
since $\mathbf{p}_1 < \mathbf{p}_2$ and

$$\lim_{s\to\infty}\frac{N-2+s}{N-1+s} = 1 \quad .$$

Further, there is a positive integer, $s_2$ , such that

$$(N-1+s_2)\mathbf{p}_1 < s_2\mathbf{p}_2 \quad .$$

The times $s_1\tau$ and $s_2\tau$ correspond to the beginning of chronological disordering (the approximate supports of the leading two particle states, $\phi_1'$ and $\phi_1''$ , overlapping significantly) and the onset of complete chronological disordering (the approximate supports of the 'first' – $\phi_1'$ – and 'last' – $\phi_1^{(N)}$ – states now overlap). In fact, this is only a very rough guide to these events, since the process of asymptotic localisation gathers progressively more of the wave-function within the velocity cone: to maintain the same tolerance on the degree of localisation a smaller velocity cone must be taken – see the next figure.



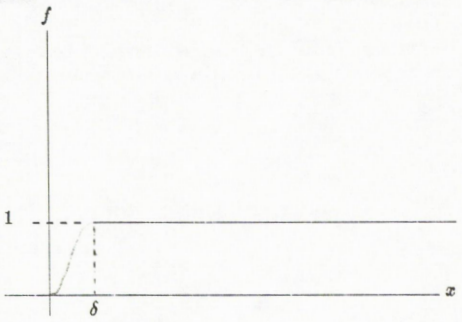

Succumbing to chronological disordering is not only hard to avoid – it is the natural condition of $N$–particle systems –, but is trivially easy; for while the component states of the beam are roughly orthonormal:

$$(i,j \in \{1,\ldots,N\}) \qquad \langle\phi_1^{(i)}|\phi_1^{(j)}\rangle \begin{cases} < \epsilon & i \neq j \\ = 1 & i = j \end{cases} ;$$

there are always observables that give a considerable overlap, i.e., where

$$\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(j)}\rangle$$

is sizable. To model a chronologically disordered system it is only necessary to take a set of observables having an 'equal' effect on all elements of the beam. For a single beam this is merely the standard (global) set of observables. For two beams there ought to be two sets of observables: one set acting on each beam and with negligible effect on the other. (The processes of chronological disordering and asymptotic localisation will not bring together beams tending to propagate in entirely opposite directions.)




It is possible to obtain observables that act on only one beam, say the $\{\phi_1^{(i)}\}$. Taking any smooth function, $f$, such that

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > \delta > 0 \end{cases} .$$

This can be used in the definition of the localising isometry, $\hat{L}_f$, discussed below (Chapter 6). The whole problem is then subjected to this isometry – the beam $\{\phi_1^{(i)}\}$ and its observables $\{\hat{A}_N\}$ –largely without changing anything; though no guarantee can be offered that this transformation is free from difficulties. However, it does indicate that there is a formal way of excluding one beam from the demesne of observables intended to act solely on another beam, and with small effect on expectation values of observations on the latter beam.

§§3.1 *Expectation values*

In expanding the expectation value for the 1–particle observable, $\hat{A}_N$, there is no need to retain the symmetrisation because the factor of $1/N$ exactly cancels the appearance of $N$ identical sums (one for $\hat{A}_1$ in each factor position). The beam is represented by the $N$–particle state

$$\Psi_N = \sum \left[\phi_1' \otimes \phi_1'' \otimes \ldots \otimes \phi_1^{(N)}\right]$$

– where $\sum[-]$ denotes the taking of an appropriately symmetrised sum. The expectation value for $\hat{A}_N = \hat{A}_1 \otimes \hat{I}_1 \otimes \ldots \otimes \hat{I}_1$ on the $N$–particle state, $\Psi_N$, is then (by an application of combinatorial analysis)

$$\begin{aligned}
&\langle\Psi_N|\hat{A}_N|\Psi_N\rangle = \\
&= \frac{1}{N!}\left(\sum_{i=1}^{N}\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(i)}\rangle(N-1)! \pm 2\Re\left(\sum_{i=1}^{N}\sum_{j=1}^{i-1}\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(j)}\rangle\langle\phi_1^{(j)}|\phi_1^{(i)}\rangle\right)(N-1)(N-2)! + \right. \\
&\qquad \left. + \text{terms of higher order in } \epsilon \right) \\
&= \frac{1}{N}\left(\sum_{i=1}^{N}\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(i)}\rangle \pm 2\Re\left(\sum_{i=1}^{N}\sum_{j=1}^{i-1}\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(j)}\rangle\langle\phi_1^{(j)}|\phi_1^{(i)}\rangle\right) + \right. \\
&\qquad \left. + \text{terms of higher order in } \epsilon \right)
\end{aligned}$$

The $(N-1)!$ and $(N-2)!$ are the number of permutations of the remaining $N-1$ and $N-2$ factors, respectively. The $N-1$ in the second sum is because the factor $\langle\phi_1^{(j)}|\phi_1^{(i)}\rangle$ may turn up in that many factor positions.

In this expression, chronological disordering terms comprehensively dominate the expectation value: the only term not due to this disordering is $\langle\phi_1'|\hat{A}_1|\phi_1'\rangle$, though in a single-beam system the

remainder of the first summation will have no adverse effect if the observable, $\hat{A}_1$ , is a constant of the motion for 1–particle states.

Where there are two beams, and a 2–particle observable $(\hat{A}_N \otimes \hat{B}_{\bar{N}})$, all the chronological disordering terms are relevant; for, in addition to the independent-pairs-of-particles terms –

$$\langle \phi_1^{(i)} | \hat{A}_1 | \phi_1^{(i)} \rangle \langle \chi_1^{(i)} | \hat{B}_1 | \chi_1^{(i)} \rangle$$

– there are also contributions –

$(i \neq j)$ $\qquad \langle \phi_1^{(i)} | \hat{A}_1 | \phi_1^{(i)} \rangle \langle \chi_1^{(j)} | \hat{B}_1 | \chi_1^{(j)} \rangle$

and

$(i \neq j)$ $\qquad \langle \phi_1^{(i)} | \hat{A}_1 | \phi_1^{(j)} \rangle \langle \chi_1^{(j)} | \hat{B}_1 | \chi_1^{(i)} \rangle$

– that represent disordered measurements.

Thus, the fact that an experiment employs a beam of the kind devised means that it cannot be conducted over an arbitrarily large distance-scale using an independent-particle hypothesis if it has been shown to work at a shorter one. The two critical times found in this section, being related to the velocity spectrum, also determine critical distances at which chronological disordering sets in and, later, becomes complete (though 'spatial disordering' might now be an apt description).

## §4 The de Broglie Paradox

The nearest that a Galilei-relativistic theory can get to a satisfactory formulation of this thought experiment is by way of the single-particle states suggested in Chapter 2, §4:

$$\Psi_1 = \frac{1}{\sqrt{2}}(\phi_1 + \chi_1) \quad . \tag{4.1}$$

The expectation of a 1–particle observable, $\hat{A}_2 = \frac{1}{2}(\hat{A}_1 \otimes \hat{I}_1 + \hat{I}_1 \otimes \hat{A}_1)$ , on a 2–particle state will be (discounting terms with factors $\langle \phi | \chi \rangle = 0$ )

$$\begin{aligned}\langle \Psi_2 | \hat{A}_2 | \Psi_2 \rangle = \frac{1}{8}\Big( & 2\langle \phi_1'' | \hat{A}_1 | \phi_1'' \rangle + 2\langle \chi_1'' | \hat{A}_1 | \chi_1'' \rangle + 2\langle \phi_1' | \hat{A}_1 | \phi_1' \rangle + 2\langle \chi_1' | \hat{A}_1 | \chi_1' \rangle + \\ & \Re\Big\{ 4\langle \chi_1'' | \hat{A}_1 | \phi_1'' \rangle + 4\langle \chi_1' | \hat{A}_1 | \phi_1' \rangle + \\ & \pm \big( 2\langle \phi_1'' | \hat{A}_1 | \phi_1' \rangle (\langle \phi_1' | \phi_1'' \rangle + \langle \chi_1' | \chi_1'' \rangle) + 2\langle \phi_1'' | \hat{A}_1 | \chi_1' \rangle (\langle \phi_1' | \phi_1'' \rangle + \langle \chi_1' | \chi_1'' \rangle) + \\ & 2\langle \chi_1'' | \hat{A}_1 | \phi_1' \rangle (\langle \phi_1' | \phi_1'' \rangle + \langle \chi_1' | \chi_1'' \rangle) + 2\langle \chi_1'' | \hat{A}_1 | \chi_1' \rangle (\langle \phi_1' | \phi_1'' \rangle + \langle \chi_1' | \chi_1'' \rangle) \big) \Big\} \Big)\end{aligned}$$

Allowing $\hat{A}_1$ to act principally on the beam $\{ \phi_1^{(i)} \}$ , leaves the radically simplified expression:

$$\langle \Psi_2 | \hat{A}_2 | \Psi_2 \rangle = \frac{1}{8}\Big( 2\langle \phi_1'' | \hat{A}_1 | \phi_1'' \rangle + 2\langle \phi_1' | \hat{A}_1 | \phi_1' \rangle \pm 2\Re\big( \langle \phi_1'' | \hat{A}_2 | \phi_1' \rangle (\langle \phi_1' | \phi_1'' \rangle + \langle \chi_1' | \chi_1'' \rangle) \big) \Big)$$

If $\hat{A}_1$ represents a constant of the motion, it is clear that chronological disordering presents a very small contribution to this expectation sum – the last term is of between first and second order in $\epsilon$ . The contribution from the half-beam not being measured – the $\{ \chi_1^{(i)} \}$ – is also very small.

In the case of an $N$–particle beam, the various terms can be summarised by considering one of each distinct form. This is done below. In considering the details of the expectation sum the only effect that symmetrisation has is to ensure that all possible combinations of 1–particle states appear on either side of the 1–particle inner-products, both with $\hat{A}_1$ and with $\hat{I}_1$ . Thus, once a certain term has been singled out it will occur with the same frequency as any other term. It is only by

introducing some sort of classification of the terms that it is possible to say that some effect, such as chronological disordering, has or does not have an appreciable impact on an observation of a beam.

There are two parts to the classification used below: first, by specifying only those factors in a term that are not $\langle\phi_1^{(i)}|\phi_1^{(i)}\rangle$ or $\langle\chi_1^{(i)}|\chi_1^{(i)}\rangle$ there are a number of ways of obtaining the form cited: for each $i$ that is not fixed there is the choice of $\phi_1^{(i)}$ or $\chi_1^{(i)}$ , i.e.,

$$\langle\phi_1^{(i)}|\phi_1^{(i)}\rangle \qquad \text{or} \qquad \langle\chi_1^{(i)}|\chi_1^{(i)}\rangle \quad ;$$

giving $2^{N-m}$ occurrences for a term in which $m$ 1–particle states are fixed (the 'Occurs' column entry).

The second part of the classification scheme consists of counting the number of ways that the superscripts can be combined for each form of term, so that $\langle\phi_1'|\hat{A}_1|\phi_1'\rangle$ is considered to be of the same 'form' as $\langle\phi_1^{(5)}|\hat{A}_1|\phi_1^{(5)}\rangle$ ; since the 'form' here is $\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(i)}\rangle$, and $i \in \{1,\ldots,N\}$ (this would give an entry of $N$ in the 'No. of Form' column). There is no extra mental effort required to deal with the fact that inner-products use states in sets of two, since if the left-hand one is fixed then for a given form of term the right-hand one is dictated by that form.

Here there are eighteen forms listed (grouped together as six 'types') – this is obviously not exhaustive, but represents the contributions roughly of order 1, $\epsilon$ and $\epsilon^2$ , beyond which I assume it is unnecessary to probe.

The normalisation constant for the sum is $(2^N N!)^{-1}$ , which is almost exactly cancelled by the product of 'Occurs' × 'No. of Form' × the number of permutations of the 1–particle expectations not involving $\hat{A}_1$ , i.e., $(N-1)!$ . The difference is independent of $N$ and is a power of 2, depending only on the 'Occurs' column.

The simplest terms are of type $a$: these are the 1–particle expectations that arise in the single-particle theory. For smallish $N$ these terms dominate the expectation sum. The second order corrections (i.e., those of order, at least, $\epsilon^2$) are not going to be considered important but are merely written down.

Terms of type $b$ are the principal contribution of chronological disordering (if $\hat{A}_1$ is not a conserved quantity there is already a large contribution from all the terms of type $a$ and $k \neq 1$ ). This part of the expectation sum is

$$\pm\Re\left(\sum_{i=1}^{N}\sum_{j=1}^{i-1}\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(j)}\rangle\left(\langle\phi_1^{(j)}|\phi_1^{(i)}\rangle + \langle\chi_1^{(j)}|\chi_1^{(i)}\rangle\right)\right)\frac{2^{N-2}(N-1)!}{2^N N!}$$

To estimate when this might become significant, the most favourable and simple assumptions will be made:

$$\langle\phi_1^{(i)}|\phi_1^{(j)}\rangle = \langle\chi_1^{(i)}|\chi_1^{(j)}\rangle = \epsilon \qquad (\forall i \neq j)$$
$$\langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(j)}\rangle = C \sim \langle\phi_1^{(i)}|\hat{A}_1|\phi_1^{(i)}\rangle \qquad (\forall i \neq j)$$

– the second assumption requires, as a minimum, that the $\{\phi_1^{(i)}\}$ share a single approximate support, as outlined in the previous sections.

The contribution of type–$b$ terms then simplifies to

$$\pm C \times N(N-1) \times 2\epsilon \times \frac{1}{4N}$$

so that this type of term may become relevant if

$$\frac{(N-1)\epsilon}{2} \approx 1 \quad .$$

So that, even if $\epsilon$ is very small, it is possible to find beams of such a size that the disordering is appreciable. There is also a crucial dependence on the observable: in the extreme case of $\hat{A}_1 \approx \hat{I}_1$ these terms are of second order in $\epsilon$ .

The last three types of term make no contribution in the usual configuration space ($\Re^3$) – after all, the observable, $\hat{A}_1$ , was chosen specifically not to apply to the $\chi$-beam. If, however, the configuration space is taken to be $S^n$ (the surface of the $n$-sphere), which is one way to model such devices as interferometers, then the two beams can separate and subsequently reconverge. The

(The use of $\pm$ indicates where symmetrising for bosons and fermions gives differing results. The indicies $i$, $j$, and $k$ are all different.)

| Form | Occurs | No. of Form | Type |
|---|---|---|---|
| $\langle\phi_1^{(k)}\|\hat{A}_1\|\phi_1^{(k)}\rangle$ | $2^{N-1}$ | $N$ | $a$ |
| $\pm\langle\phi_1^{(k)}\|\hat{A}_1\|\phi_1^{(k)}\rangle\langle\phi_1^{(i)}\|\phi_1^{(j)}\rangle\langle\phi_1^{(j)}\|\phi_1^{(i)}\rangle$ | $2^{N-3}$ | $N(N-1)(N-\hat{}$ | $a$ |
| $\pm\langle\phi_1^{(k)}\|\hat{A}_1\|\phi_1^{(k)}\rangle\Re\left(\langle\phi_1^{(i)}\|\phi_1^{(j)}\rangle\langle\chi_1^{(i)}\|\chi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)(N-2)$ | $a$ |
| $\pm\langle\phi_1^{(k)}\|\hat{A}_1\|\phi_1^{(k)}\rangle\langle\chi_1^{(i)}\|\chi_1^{(j)}\rangle\langle\chi_1^{(j)}\|\chi_1^{(i)}\rangle$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $a$ |
| $\pm\Re\left(\langle\phi_1^{(k)}\|\hat{A}_1\|\phi_1^{(l)}\rangle\langle\phi_1^{(l)}\|\phi_1^{(k)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $b$ |
| $\pm\Re\left(\langle\phi_1^{(k)}\|\hat{A}_1\|\phi_1^{(l)}\rangle\langle\chi_1^{(l)}\|\chi_1^{(k)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $b$ |
| 'Re-interference' | | | |
| $2\Re\left(\langle\phi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\right)$ | $2^{N-1}$ | $N$ | $c$ |
| $\pm\langle\phi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\langle\phi_1^{(i)}\|\phi_1^{(j)}\rangle\langle\phi_1^{(j)}\|\phi_1^{(i)}\rangle$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $c$ |
| $\pm\Re\left(\langle\phi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\langle\phi_1^{(i)}\|\phi_1^{(j)}\rangle\langle\chi_1^{(i)}\|\chi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)(N-2)$ | $c$ |
| $\pm\langle\phi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\langle\chi_1^{(i)}\|\chi_1^{(j)}\rangle\langle\chi_1^{(j)}\|\chi_1^{(i)}\rangle$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $c$ |
| $\pm2\Re\left(\langle\phi_1^{(i)}\|\hat{A}_1\|\chi_1^{(j)}\rangle\langle\phi_1^{(j)}\|\phi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $d$ |
| $\pm2\Re\left(\langle\phi_1^{(i)}\|\hat{A}_1\|\chi_1^{(j)}\rangle\langle\chi_1^{(j)}\|\chi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $d$ |
| $\langle\chi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle$ | $2^{N-1}$ | $N$ | $e$ |
| $\pm\langle\chi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\langle\phi_1^{(i)}\|\phi_1^{(j)}\rangle\langle\phi_1^{(j)}\|\phi_1^{(i)}\rangle$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $e$ |
| $\pm\Re\left(\langle\chi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\langle\phi_1^{(i)}\|\phi_1^{(j)}\rangle\langle\chi_1^{(i)}\|\chi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)(N-2)$ | $e$ |
| $\pm\langle\chi_1^{(k)}\|\hat{A}_1\|\chi_1^{(k)}\rangle\langle\chi_1^{(i)}\|\chi_1^{(j)}\rangle\langle\chi_1^{(j)}\|\chi_1^{(i)}\rangle$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $e$ |
| $\pm\Re\left(\langle\chi_1^{(i)}\|\hat{A}_1\|\chi_1^{(j)}\rangle\langle\phi_1^{(j)}\|\phi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $f$ |
| $\pm\Re\left(\langle\chi_1^{(i)}\|\hat{A}_1\|\chi_1^{(j)}\rangle\langle\chi_1^{(j)}\|\chi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $f$ |

interference patterns obtained are of some contemporary interest in, for example, neutron beam studies. This interference is supposed to arise from the overlap term, $\langle\phi_1'|\hat{A}_1|\chi_1'\rangle$ . Clearly, if $\hat{A}_1$ is not a conserved quantity the other terms of type $c$ (i.e., for $k \neq 1$ ) will modify this. In the case of an interferometer, $\hat{A}_1 = E(\hat{\mathbf{x}}; \Lambda_M)$ , and the fringes for one particle at any instant are unlikely to match up with those of any other particle in the beam, which will tend to smooth out the maxima and minima, and so blur the pattern of fringes. There is no danger of $\hat{A}_1$ approaching $\hat{I}_1$ here, because interference patterns are captured on photographic plates (or something equivalent to this), which cover one fewer dimension than the space in which the states have their support: for $S^3$ , the plate is two-dimensional; on $S^1$ , the interference 'pattern' is the value of an observable on a single point.

Again, if

$$N\epsilon \approx 1$$

there are all the terms of type $d$ and $f$ to take into account.

No attempt has been made here to consider the case of an $N$–particle beam in which there are initially large overlaps between the 1–particle states – this would be the result of taking $\epsilon \approx 1$ .

Likewise, no consideration has been given to the possible disturbance of the beam – something that is more likely the further the beam travels – indeed, in the case of the $S^3$ configuration space, e.g., an interferometer, there is obviously a disturbance that has been assumed to be exactly symmetric between the two half-beams.

## §5 Bohm's revision of the Einstein–Rosen–Podolsky paradox

For some time now it has seemed incredible to me that Bohm and Aharonov's revision of the Einstein–Rosen–Podolsky (BERP) paradox[1] could present a devastating contradiction between quantum mechanics and Einstein's postulate of special relativity. The paradox is purported to arise from an action-at-a-distance that connects two space-like separated measurements. The peculiarity is that Bohm's theory at no point contains any reference to space or time. It is explained that some pair of particles separates in two different directions and is later measured, but nowhere in the formulation of the model is this statement incorporated: the state vector is taken as

$$\Psi_2 = \frac{1}{\sqrt{2}}(\alpha \otimes \beta - \beta \otimes \alpha)$$

where $\alpha$ represents spin-up and $\beta$ spin-down in the $z$-direction.

Bohm's mathematics must therefore be regarded as only a loose guide to the experimental realisations of his explanation. In fact, Bohm's model applies without modification to the case of two spin-correlated particles travelling together in such a way that they cannot be distinguished; a rather uncontroversial situation in which to find correlations.

A somewhat more specific formulation of the paradox is, therefore, most in order. Framing a Galilei-relativistic theory has precisely the same pitfalls as covered above for the de Broglie paradox; and the same compromise with accuracy is necessary to get around these difficulties. That aside, there is no reason to expect any radical revision of Bohm's result.

To formulate the de Broglie paradox two sets of single-particle states were mixed to form the beam, here four sets will be required (using the conditions of §3):

| State | Momentum | $z$ spin |
|---|---|---|
| $\phi_1^{(i)}$ | $[\mathbf{p}_1, \mathbf{p}_2]$ | spin-up |
| $\psi_1^{(i)}$ | | spin-down |
| $\chi_1^{(i)}$ | $[-\mathbf{p}_2, -\mathbf{p}_1]$ | spin-down |
| $\xi_1^{(i)}$ | | spin-up |

The state vector for a pair of particles in the singlet state is then

$$\Psi_2 = \frac{1}{2}(\phi_1 \otimes \chi_1 - \psi_1 \otimes \xi_1 \pm [\chi_1 \otimes \phi_1 - \xi_1 \otimes \psi_1])$$

(+ $\Rightarrow$ bosons; − $\Rightarrow$ fermions). A 2-particle observable is required:

$$\hat{A}_2 = \frac{1}{2}\left(\hat{A}_1 \otimes \hat{B}_1 + \hat{B}_1 \otimes \hat{A}_1\right) \quad .$$

As before, $\hat{A}_1$ and $\hat{B}_1$ will be taken to act only on the approximately localised particles moving in one direction: $\hat{A}_1$ acting on $\phi_1$ and $\psi_1$ , $\hat{B}_1$ acting on $\chi_1$ and $\xi_1$ . The expectation value is then,

in full,

$$\begin{aligned}\langle\Psi_2|\hat{A}_2|\Psi_2\rangle = \frac{1}{4}\Big(&\langle\phi_1|\hat{A}_1|\phi_1\rangle\langle\chi_1|\hat{B}_1|\chi_1\rangle + \langle\psi_1|\hat{A}_1|\psi_1\rangle\langle\xi_1|\hat{B}_1|\xi_1\rangle + \langle\chi_1|\hat{A}_1|\chi_1\rangle\langle\psi_1|\hat{B}_1|\phi_1\rangle \\ &+ \langle\xi_1|\hat{A}_1|\xi_1\rangle\langle\psi_1|\hat{B}_1|\psi_1\rangle + 2\Re\big(-\langle\phi_1|\hat{A}_1|\psi_1\rangle\langle\chi_1|\hat{B}_1|\xi_1\rangle \pm \langle\phi_1|\hat{A}_1|\xi_1\rangle\langle\chi_1|\hat{B}_1|\psi_1\rangle \\ &\pm \langle\psi_1|\hat{A}_1|\chi_1\rangle\langle\xi_1|\hat{B}_1|\phi_1\rangle \pm \langle\psi_1|\hat{A}_1|\xi_1\rangle\langle\xi_1|\hat{B}_1|\psi_1\rangle - \langle\chi_1|\hat{A}_1|\xi_1\rangle\langle\phi_1|\hat{B}_1|\psi_1\rangle \\ &\pm \langle\phi_1|\hat{A}_1|\chi_1\rangle\langle\chi_1|\hat{B}_1|\phi_1\rangle\big)\Big)\end{aligned}$$

Taking into account the null spaces of the observables:

$$\langle\Psi_2|\hat{A}_2|\Psi_2\rangle = \frac{1}{4}\Big(\langle\phi_1|\hat{A}_1|\phi_1\rangle\langle\chi_1|\hat{B}_1|\chi_1\rangle - 2\Re\big(\langle\phi_1|\hat{A}_1|\psi_1\rangle\langle\chi_1|\hat{B}_1|\xi_1\rangle\big) + \langle\psi_1|\hat{A}_1|\psi_1\rangle\langle\xi_1|\hat{B}_1|\xi_1\rangle\Big)$$

The right-hand side of this is half what it should be, for if $\hat{A}_2$ is taken to be the product of identity operators for the two beam 'directions':

$$\hat{A}_2 = E(\hat{p};[0,\infty]) \otimes E(\hat{p};[-\infty,0])$$

the result is 1/2 not 1. This factor of 1/2 is a consequence of taking $\Psi_2$ to have a fully symmetrised form, whereas the class of observables distinguishes between two disjoint sub-spaces of states. Since they are distinguishable, there is no need to symmetrise states with momentum spectra of opposing signs. The appropriate state to use is, therefore,

$$\Psi_2' = \frac{1}{\sqrt{2}}(\phi_1 \otimes \chi_1 - \psi_1 \otimes \xi_1)$$

If $\hat{A}_1$ and $\hat{B}_1$ are now taken to be measurements of spin, $\hat{\sigma}^1$ and $\hat{\sigma}^2$ , and if the states are re-written in terms of their spin-dependence in the conventional manner:

$$\langle\Psi_2|\hat{A}_2|\Psi_2\rangle = \frac{1}{2}\Big(\langle\alpha|\hat{\sigma}^1|\alpha\rangle\langle\beta|\hat{\sigma}^2|\beta\rangle - 2\Re\big(\langle\alpha|\hat{\sigma}^1|\beta\rangle\langle\beta|\hat{\sigma}^2|\alpha\rangle\big) + \langle\beta|\hat{\sigma}^1|\beta\rangle\langle\alpha|\hat{\sigma}^2|\alpha\rangle\Big)$$

which is exactly Bohm's result. This is given its more familiar form by writing

$$\hat{\sigma}^i = \hat{\sigma}_z \cos\theta_i + \hat{\sigma}_x \sin\theta_i \quad ,$$

and, since

$$\begin{aligned}\langle\alpha|\hat{\sigma}_z|\alpha\rangle &= 1 & \langle\alpha|\hat{\sigma}_x|\alpha\rangle &= 0 \\ \langle\alpha|\hat{\sigma}_z|\beta\rangle &= 0 & \langle\alpha|\hat{\sigma}_x|\beta\rangle &= 1 \\ \langle\beta|\hat{\sigma}_z|\beta\rangle &= -1 & \langle\beta|\hat{\sigma}_x|\beta\rangle &= 0\end{aligned}$$

it follows that

$$\begin{aligned}\langle\alpha|\hat{\sigma}^1|\alpha\rangle\langle\beta|\hat{\sigma}^2|\beta\rangle &= \cos\theta_1(-\cos\theta_2) \\ \langle\alpha|\hat{\sigma}^2|\alpha\rangle\langle\beta|\hat{\sigma}^1|\beta\rangle &= \cos\theta_2(-\cos\theta_1) \\ \langle\alpha|\hat{\sigma}^1|\beta\rangle\langle\beta|\hat{\sigma}^2|\alpha\rangle &= \sin\theta_1 \sin\theta_2\end{aligned}$$

so

$$\begin{aligned}\langle\Psi_2|\hat{A}_2|\Psi_2\rangle &= -\frac{1}{2}(2\cos\theta_2\cos\theta_1 + 2\sin\theta_1\sin\theta_2) \\ &= -\cos(\theta_1 - \theta_2) \quad .\end{aligned}$$

Consider, now, two pairs of particles

$$\Psi_4 = \sum[(\phi_1' \otimes \chi_1' - \psi_1' \otimes \xi_1') \otimes (\phi_1'' \otimes \chi_1'' - \psi_1'' \otimes \xi_1'')]$$

where $\sum[-]$ means the appropriately symmetrised and normalized sum. In fact, only states going in the same direction need be symmetrised. By the same argument used in the last section, the observable need not be symmetrised when computing an expectation value because this only produces a number of identical expressions divided by that number. Assuming that the first and third factors

represent one direction, and the second and fourth factors represent the opposite direction of motion, the 2-particle observable used is

$$\hat{A}_4 = \hat{A}_1 \otimes \hat{B}_1 \otimes \hat{I}_1 \otimes \hat{I}_1 \quad .$$

Ignoring terms involving powers of $\epsilon$ above one, the expectation sum is

$$\begin{aligned}
&\langle\Psi_4|\hat{A}_4|\Psi_4\rangle = \\
&\quad \frac{1}{16}\Big(\langle\phi_1'|\hat{A}_1|\phi_1'\rangle\left[2\langle\chi_1'|\hat{B}_1|\chi_1'\rangle + \langle\chi_1''|\hat{B}_1|\chi_1''\rangle + \langle\xi_1''|\hat{B}_1|\xi_1''\rangle \pm 2\Re(\langle\chi_1'|\hat{B}_1|\chi_1''\rangle\langle\chi_1''|\chi_1'\rangle)\right] + \\
&\qquad \langle\phi_1''|\hat{A}_1|\phi_1''\rangle\left[2\langle\chi_1''|\hat{B}_1|\chi_1''\rangle + \langle\chi_1'|\hat{B}_1|\chi_1'\rangle + \langle\xi_1'|\hat{B}_1|\xi_1'\rangle \pm 2\Re(\langle\chi_1''|\hat{B}_1|\chi_1'\rangle\langle\chi_1'|\chi_1''\rangle)\right] + \\
&\qquad \langle\psi_1'|\hat{A}_1|\psi_1'\rangle\left[2\langle\xi_1'|\hat{B}_1|\xi_1'\rangle + \langle\xi_1''|\hat{B}_1|\xi_1''\rangle + \langle\chi_1''|\hat{B}_1|\chi_1''\rangle \pm 2\Re(\langle\xi_1'|\hat{B}_1|\xi_1''\rangle\langle\xi_1''|\xi_1'\rangle)\right] + \\
&\qquad \langle\psi_1''|\hat{A}_1|\psi_1''\rangle\left[2\langle\xi_1''|\hat{B}_1|\xi_1''\rangle + \langle\xi_1'|\hat{B}_1|\xi_1'\rangle + \langle\chi_1'|\hat{B}_1|\chi_1'\rangle \pm 2\Re(\langle\xi_1''|\hat{B}_1|\xi_1'\rangle\langle\xi_1'|\xi_1''\rangle)\right] + \\
&\qquad -4\Re\left(\langle\phi_1''|\hat{A}_1|\psi_1''\rangle\langle\chi_1''|\hat{B}_1|\xi_1''\rangle + \langle\phi_1'|\hat{A}_1|\psi_1'\rangle\langle\chi_1'|\hat{B}_1|\xi_1'\rangle\right) \\
&\qquad + \text{more terms of order } \epsilon \text{ , or worse.}\Big)
\end{aligned}$$

The result is half of Bohm's, though a condition has to be added that $\epsilon$ be small for this to be so.

The table of terms for a pair of $N$–particle beams is not dissimilar to the de Broglie case: the extra terms arising because where before there was a 1–particle state there are now two to take into account.

| Form | Occurs | No. of Form | Type |
|---|---|---|---|
| $\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(i)}\rangle\langle\chi_1^{(i)}\|\hat{B}_1\|\chi_1^{(i)}\rangle$ | $2^{N-1}$ | $N$ | $a$ |
| $\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(i)}\rangle\langle\chi_1^{(j)}\|\hat{B}_1\|\chi_1^{(j)}\rangle$ | $2^{N-2}$ | $N(N-1)$ | $b$ |
| $\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(i)}\rangle\langle\xi_1^{(j)}\|\hat{B}_1\|\xi_1^{(j)}\rangle$ | $2^{N-2}$ | $N(N-1)$ | $b'$ |
| $\pm\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(i)}\rangle\Re\left(\langle\chi_1^{(i)}\|\hat{B}_1\|\chi_1^{(j)}\rangle\langle\chi_1^{(j)}\|\chi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $c$ |
| $\pm\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(i)}\rangle\Re\left(\langle\chi_1^{(j)}\|\hat{B}_1\|\chi_1^{(k)}\rangle\langle\chi_1^{(k)}\|\chi_1^{(j)}\rangle\right)$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $d$ |
| $\pm\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(i)}\rangle\Re\left(\langle\xi_1^{(j)}\|\hat{B}_1\|\xi_1^{(k)}\rangle\langle\xi_1^{(k)}\|\xi_1^{(j)}\rangle\right)$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $d'$ |
| $\pm\langle\chi_1^{(i)}\|\hat{B}_1\|\chi_1^{(i)}\rangle\Re\left(\langle\phi_1^{(i)}\|\hat{A}_1\|\phi_1^{(j)}\rangle\langle\phi_1^{(j)}\|\phi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $c$ |
| $\pm\langle\chi_1^{(i)}\|\hat{B}_1\|\chi_1^{(i)}\rangle\Re\left(\langle\phi_1^{(j)}\|\hat{A}_1\|\phi_1^{(k)}\rangle\langle\phi_1^{(k)}\|\phi_1^{(j)}\rangle\right)$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $d$ |
| $\pm\langle\chi_1^{(i)}\|\hat{B}_1\|\chi_1^{(i)}\rangle\Re\left(\langle\psi_1^{(j)}\|\hat{A}_1\|\psi_1^{(k)}\rangle\langle\psi_1^{(k)}\|\psi_1^{(j)}\rangle\right)$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $d'$ |
| $\langle\psi_1^{(i)}\|\hat{A}_1\|\psi_1^{(i)}\rangle\langle\xi_1^{(i)}\|\hat{B}_1\|\xi_1^{(i)}\rangle$ | $2^{N-1}$ | $N$ | $a'$ |
| $\langle\psi_1^{(i)}\|\hat{A}_1\|\psi_1^{(i)}\rangle\langle\chi_1^{(j)}\|\hat{B}_1\|\chi_1^{(j)}\rangle$ | $2^{N-2}$ | $N(N-1)$ | $b$ |
| $\langle\psi_1^{(i)}\|\hat{A}_1\|\psi_1^{(i)}\rangle\langle\xi_1^{(j)}\|\hat{B}_1\|\xi_1^{(j)}\rangle$ | $2^{N-2}$ | $N(N-1)$ | $b'$ |
| $\pm\langle\psi_1^{(i)}\|\hat{A}_1\|\psi_1^{(i)}\rangle\Re\left(\langle\xi_1^{(i)}\|\hat{B}_1\|\xi_1^{(j)}\rangle\langle\xi_1^{(j)}\|\xi_1^{(i)}\rangle\right)$ | $2^{N-2}$ | $N(N-1)$ | $c'$ |
| $\pm\langle\psi_1^{(i)}\|\hat{A}_1\|\psi_1^{(i)}\rangle\Re\left(\langle\xi_1^{(j)}\|\hat{B}_1\|\xi_1^{(k)}\rangle\langle\xi_1^{(k)}\|\xi_1^{(j)}\rangle\right)$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $d$ |
| $\pm\langle\psi_1^{(i)}\|\hat{A}_1\|\psi_1^{(i)}\rangle\Re\left(\langle\chi_1^{(j)}\|\hat{B}_1\|\chi_1^{(k)}\rangle\langle\chi_1^{(k)}\|\chi_1^{(j)}\rangle\right)$ | $2^{N-3}$ | $N(N-1)(N-2)$ | $d'$ |

$$\pm\langle\xi_1^{(i)}|\hat{B}_1|\xi_1^{(i)}\rangle\Re\left(\langle\psi_1^{(i)}|\hat{A}_1|\psi_1^{(j)}\rangle\langle\psi_1^{(j)}|\psi_1^{(i)}\rangle\right) \qquad 2^{N-2} \qquad N(N-1) \qquad c'$$

$$\pm\langle\xi_1^{(i)}|\hat{B}_1|\xi_1^{(i)}\rangle\Re\left(\langle\psi_1^{(j)}|\hat{A}_1|\psi_1^{(k)}\rangle\langle\psi_1^{(k)}|\psi_1^{(j)}\rangle\right) \qquad 2^{N-3} \qquad N(N-1)(N-2) \qquad d$$

$$\pm\langle\xi_1^{(i)}|\hat{B}_1|\xi_1^{(i)}\rangle\Re\left(\langle\phi_1^{(j)}|\hat{A}_1|\phi_1^{(k)}\rangle\langle\phi_1^{(k)}|\phi_1^{(j)}\rangle\right) \qquad 2^{N-3} \qquad N(N-1)(N-2) \qquad d'$$

$$-2\Re\left(\langle\phi_1^{(i)}|\hat{A}_1|\psi_1^{(i)}\rangle\langle\chi_1^{(i)}|\hat{B}_1|\xi_1^{(i)}\rangle\right) \qquad 2^{N-1} \qquad N \qquad e$$

$$\mp2\Re\left(\langle\phi_1^{(i)}|\hat{A}_1|\psi_1^{(i)}\rangle\langle\chi_1^{(i)}|\hat{B}_1|\xi_1^{(j)}\rangle\langle\xi_1^{(j)}|\xi_1^{(i)}\rangle\right) \qquad 2^{N-2} \qquad N(N-1) \qquad f$$

$$\mp2\Re\left(\langle\phi_1^{(i)}|\hat{A}_1|\psi_1^{(i)}\rangle\langle\chi_1^{(j)}|\hat{B}_1|\xi_1^{(i)}\rangle\langle\chi_1^{(i)}|\chi_1^{(j)}\rangle\right) \qquad 2^{N-2} \qquad N(N-1) \qquad f$$

$$\mp2\Re\left(\langle\chi_1^{(i)}|\hat{B}_1|\xi_1^{(i)}\rangle\langle\phi_1^{(i)}|\hat{A}_1|\psi_1^{(j)}\rangle\langle\psi_1^{(j)}|\psi_1^{(i)}\rangle\right) \qquad 2^{N-2} \qquad N(N-1) \qquad f$$

$$\mp2\Re\left(\langle\chi_1^{(i)}|\hat{B}_1|\xi_1^{(i)}\rangle\langle\phi_1^{(j)}|\hat{A}_1|\psi_1^{(i)}\rangle\langle\phi_1^{(i)}|\phi_1^{(j)}\rangle\right) \qquad 2^{N-2} \qquad N(N-1) \qquad f$$

The normalisation factor for the expectation sum is

$$\frac{1}{(N!)^2}\frac{1}{(\sqrt{2})^{2N}}$$

– the product of symmetrising the state and the $1/\sqrt{2}$ each singlet-state introduces.

The multiplicity that each form contributes to the expectation is calculated by

$$\text{'Occurs'} \times \text{'No. of Form'} \times \left((N-1)!\right)^2$$

– the $\left((N-1)!\right)^2$ comes from the permutation of factors of the form $\langle\eta|\eta\rangle$ .

In the BERP case, if I make the simplistic assumption that the beams are separated but otherwise the spatial distribution can be ignored, the prospects for chronological disordering making a substantial impact are diminished by large-scale cancellations. Term type $b$ cancels term type $b'$; likewise for $d$ and $d'$. There is, therefore, no mixing of pairs if $\epsilon$ is small.

The remainder of the table can be written out again in the notation of the BERP observables:

| Form | Multiplicity | Type |
|---|---|---|
| $-\cos\theta_1\cos\theta_2$ | $2^{N-1}N\left((N-1)!\right)^2$ | $a$ |
| $(\cos\theta_1\cos\theta_2)\epsilon$ | $2\times2^{N-2}N(N-1)\left((N-1)!\right)^2$ | $c$ |
| $-\cos\theta_1\cos\theta_2$ | $2^{N-1}N\left((N-1)!\right)^2$ | $a'$ |
| $(\cos\theta_1\cos\theta_2)\epsilon$ | $2\times2^{N-2}N(N-1)\left((N-1)!\right)^2$ | $c'$ |
| $-2\sin\theta_1\sin\theta_2$ | $2^{N-1}N\left((N-1)!\right)^2$ | $e$ |
| $(2\sin\theta_1\sin\theta_2)\epsilon$ | $4\times2^{N-2}N(N-1)\left((N-1)!\right)^2$ | $f$ |

Bohm's correlation function now only turns up if $\epsilon$ is ignorable. The first two terms in an expansion in powers of $\epsilon$ are:

$$1: \qquad -\cos(\theta_1-\theta_2)\frac{2\times2^{N-1}\times N\times\left((N-1)!\right)^2}{2^N\times(N!)^2} = -\cos(\theta_1-\theta_2)\frac{1}{N}$$

$$\epsilon: \qquad (2\sin\theta_1\sin\theta_2+\cos\theta_1\cos\theta_2)\frac{2\times2^N\times N(N-1)\times\left((N-1)!\right)^2}{2^N\times(N!)^2}$$

$$= (2\sin\theta_1\sin\theta_2+\cos\theta_1\cos\theta_2)\frac{2(N-1)}{N}\epsilon$$

The $\epsilon^2$ contribution will result from terms similar to $a$ and $e$, but with two fewer factors of unity.

Accepting the pretty drastic simplifications employed, the conclusion here is that the correlation decays with $1/N$ , until it reaches about $\epsilon\cos(\theta_1 - \theta_2)$ , where the decrease levels out. Rather perversely, it is the chronological disordering terms that avert the complete disappearance of correlations, albeit at a very low level. It seems unlikely that any more realistic calculation will increase the correlation.

What all this means is the quantum mechanical correlations that form part of the statement of Bell's theorem are not always of the form deduced by Bohm. In the case of a beam of sufficiently many particles that are chronologically disordered, Bell's equation 22 becomes[2]

$$4(\epsilon + \delta) \geq \frac{|\vec{a}\cdot\vec{c} - \vec{a}\cdot\vec{b}|}{N} + \frac{\vec{b}\cdot\vec{c}}{N} - 1 \quad .$$

So that the lower bound on $\epsilon + \delta$ is about $-1/4$. It would be an exaggeration to say this resolved the paradox; after all, an experiment can always be adjusted to use beams of fewer particles. If this shows anything it is that a quantum mechanical effect disappears under certain circumstances – more particles, or longer beam pipes.

## §6 Conclusion

After a certain amount of work to accommodate the limitations of a Galilei-relativistic approach, a multi-particle theory for various free evolutions has been produced. In this context there is a real phenomenon that has been given the name *chronological disordering*.

The first criterion for chronological disordering to be significant arises from the assumption that, initially, the particles in a beam are clearly distinguishable. The Asymptotic Localisation Theorem leads to the existence of a time (or distance) beyond which this *independent-particle hypothesis* is no longer tenable.

A second criterion is based on the degree to which the different states are orthogonal. For a particular beam this near-orthogonality is overcome by taking a sufficient number of particles in each beam – after the first criterion has been met.

In the case of the de Broglie paradox, this would seem to lead to a blurring of interference fringes when, for example, a beam of neutrons is split round the arms of an interferometer. So, if a clear pattern is obtained with an interferometer that has path lengths of the order of metres, there is every reason to think that this will not be the case if the path lengths are of the order of kilometres.

In the case of Bohm's revision of the Einstein–Rosen–Podolsky paradox, the correlation function is a multiple of the single-particle case. The 2–particle observable clearly registers some effect when the independent–particle hypothesis fails: the correlation diminishes at a rate proportional to the number of pairs in the chronologically disordered system, i.e., the number of indistinguishable particles in either beam that are in the vicinity of each spin detector.

The results presented here are rather less than precise: it is not at all clear when chronological disordering – if such a miscellany of different terms in expectation sums can be usefully collected under any single heading – sets in, and to what extent. There is precious little but guess-work in the selection of states, localisations, observables and $\epsilon$ . It is far from clear that a small change in any one of these guesses will not produce a large change in the expectation value. The contrast with the precision of atomic and molecular modelling could not be sharper.

The conclusions reached here are (I would say, 'of necessity') vague, if suggestive. General discussions always face this pitfall; and a specific example is the best cure. If this were felt to be worthwhile, the next step in pursuit of this topic would be to represent some actual experiment in the manner presented here. One of the principal objections to this whole approach is the reliance upon the Galilei symmetry group: if spatial separations between beams are to be meaningfully modelled, it might be more natural to use a Poincaré-relativistic formulation.

## §7 Bibliography

The concept of chronological disordering first appears in a primitive form in

K. K. Wan, D. R. E. Timson, *Conservation laws in asymptotically separable quantum mechanics*, Phys. Lett. **111a** (1985) 165.

and

D. R. E. Timson, *Locality in non-relativistic and relativistic quantum mechanics*, Ph. D. Thesis, University of St. Andrews, 1986; pp 102–133.

where the arguments used and the conclusions drawn differ almost entirely from those found here.

10: [1] David Bohm, Y. Aharonov, *Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky*, Phys. Rev. **108** (1957) 1070–1076.

14: [2] J. S. Bell, *On the Einstein Podolsky Rosen Paradox*, Physics **1** (1964) 195–200.

# Chapter 4

# A Survey of Position and Poincaré-Invariance in Quantum Mechanics

*a perfidious history*

*'We are the music makers,*
*We are the dreamers of dreams,*
*Wandering by lonely sea-breakers,*
*And sitting by desolate streams;–*
*World-losers and world-forsakers,*
*On whom the pale moon gleams:*
*We are the movers and shakers*
*Of the world for ever, it seems.'*

Arthur William Edgar O'Shaughnessy, **Ode: We are the music makers.**

## §1 Prelude

Before introducing my survey it is essential to point out that this will be, by no means, a complete review. To my knowledge only Andrés Kálnay has attempted that Sisyphean task.[1] Rather than try to up-date Kálnay's feat of superhuman distillation and eccentric English, I have adopted a far more selective strategy. After all, as Kálnay ironically remarked, 'Much work [has been] done on position and velocity in relativistic quantum mechanics: see Refs. [1–87].' The field has continued to flourish since then; or is it just in more desperate need of weeding?

Since the purpose of this survey was to find a viable beginning for a more credible theory, I have ruthlessly ignored a host of authors; those lucky enough to have found a place here are certainly not accorded the respect and consideration they deserve. The ideas analysed may not be classed as adequate but their careful development has been an indispensible milieu for my own, rather presumptuous, notions.

An outline of the survey will occupy the remainder of this section.

Until a satisfactory solution to the problem of locality attains general acceptance, the centre of the debate will continue to be the 1949 paper by T. D. Newton and Eugene Wigner. For despite self-confessed failure, this remains the best argued conception. Arthur Wightman has even gone so far as to renounce Newton and Wigner's admission of error.[2]

To put the issues into context, I have chosen to place the beginning of my history a little earlier than 1949; for in 1935 Maurice Pryce was if not the first then certainly one of the earlier authors to ponder the definition of a 'position observable' – his paper immediately follows one in which Born and Infeld had introduced an 'energy centre'. By 1948 Pryce was attempting a comprehensive listing of all possibilities. Several of Pryce's conclusions are also found in the work of Anastasios Papapetrou and Christian Møller. The earliest work, of Erwin Schrödinger (Berl. Ber. (1930) 418; (1931) 63), I have eschewed on linguistic grounds. These early ruminations are based upon the constructs of the mechanics of continuous media: centres of mass, inertia and gravity being obtained from the appropriate density field. There is thus an attempt to repeat the Newtonian centres through which forces can be said to act in a context where such centres, and the rigid bodies they presuppose, are not valid.

By contrast, Newton and Wigner treated the problem as an application of 'axiomatics': formulating a set of symmetry requirements on a space of states that produced a unique operator for each value of spin. Asim Barut and S. Malin analysed this approach, only to conclude that it did not satisfy all reasonable symmetry criteria (i.e., Lorentz invariance) – just as Newton and Wigner had tersely admitted. This was followed by Gordon Fleming's ploy to circumvent the lack of invariance by making explicit in his formulation the constant-time hyperplane on which measurements are to be based. This does not succeed because the implied assumption is either that there is a preferred frame of reference or that each frame of reference has its own set of observables, which cannot thereafter be related to observables in any other frame.

Several authors have tried using a 'proper time' as the evolution parameter (Collins, Fanchi, Cook, Horwitz, Piron, Reuse, Vigier *et al.*, Nambu, and Fock), often by the production of a Schrödinger-type wave equation of first order in the 'proper time' coordinate. Such efforts are confounded by difficulties over normalisation, the definition of observables, and the meaning of this 'proper time'. Often the 'proper time' is more of a 'super time' (as Olivier Costa de Beauregard has put it[3]) which produces a different wave function over space-time for each separate 'super time' instant.

During the 1960's, Joachim Petzold, Bernd Gerlach and Dieter Gromes attacked from another direction. They tried to devise a 4-vector probability current with satisfactory symmetry and 'causality' (development in space-time) properties, having first shown that the commonly-quoted 4-currents were not satisfactory. In Appendix B of this Thesis I have obtained a translation of one of their key papers. The most direct criticism of this work is that, instead of finding a 4-current, Petzold, Gerlach and Gromes have found a countable infinity of candidates. In the absence of any compelling, additional criteria, this result can only be regarded with suspicion.

The whole range of physics journals is littered with the hopeful beginnings of different Poincaré-invariant quantum theories. Theorists in search of concrete results have long-since moved on to field theories, where the issue of Position either does not appear or may be ignored. Given the fifty-some years that the problem has festered, who can blame them?

## §2 Early Fumblings: 1935–49

M. H. L. Pryce, *Commuting Co-ordinates in the New Field Theory*, Proc. Roy. Soc. **150A** (1935) 166–172.
M. H. L. Pryce, *The mass-centre in the restricted theory of relativity and its connexion with the quantum theory of elementary particles*, Proc. Roy. Soc. **195A** (1948) 62–81.
A. Papapetrou, *The concepts of angular momentum and the centre of gravity in relativity mechanics*, Praktika Akad. Athenon **14** (1939) 540–547. (See Appendix A for a translation of this paper.)
Chr. Møller, *On the definition of the centre of gravity of an arbitrary closed system in the theory of relativity*, Comm. Dublin Inst. Adv. Stud. ser. A no. 5 (1949).
Chr. Møller, *The Theory of Relativity*, O. U. P. 1972; pp 132–133, and pp 144–187.

The position observable in a Galilei-invariant quantum theory is a simple, straightforward quantity that has never been questioned. It must have come as something of a shock to the theorists who tried to progress onto a Poincaré-invariant theory that this elementary concept could not be formulated. The authors listed above attacked the problem from the mechanics of continuous media: the 'position' is an a average over the matter distribution that gives some 'centre' of this distribution – a centre of 'gravity', 'mass', or 'inertia'. The distribution of the system under study is given by the (symmetric) stress-energy tensor, $T^{\mu\nu}$ .

The total momentum of the system is defined to be

$$P^\mu = \int T^{0\mu}\, d^3\mathbf{x} \quad . \tag{2.1}$$

The total angular momentum is defined to be (the spatial part of)

$$M^{\mu\nu} = \int x^\mu T^{0\nu} - x^\nu T^{0\mu}\, d^3\mathbf{x} \quad . \tag{2.2}$$

The laws of conservation of energy, momentum, and angular momentum are ensured by the condition

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu} = 0 \quad , \tag{2.3}$$

and the requirement that $T^{\mu\nu}$ is a regular function of finite spatial support.

The momentum-energy conservation law is deduced by setting $\mu$ in equation 3 and integrating over a constant-time hyperplane: the space terms ( $\nu = 1, 2, 3$ ) vanish because $T^{\mu\nu}$ is bounded, leaving

$$\frac{\partial}{\partial x^0} \int T^{\mu 0}\, d^3\mathbf{x} = 0$$

i.e.,

$$\frac{\partial P^\mu}{\partial x^0} = 0 \quad .$$

It follows from equation 3 and the symmetry of $T^{\mu\nu}$ that

$$\frac{\partial}{\partial x^\nu}\left(x^\mu T^{\sigma\nu} - x^\sigma T^{\mu\nu}\right) = 0 \quad ;$$

applying the same argument as just above then gives

$$\frac{\partial M^{\mu\nu}}{\partial x^0} = 0 \quad .$$

To show that these conserved quantities represent physical attributes of the total system, it is necessary to show that $P$ is a 4–vector and $M$ is a second-rank tensor. The method is essentially the same in both cases, so only the proof for $P$ will be given.

Take an arbitrary but constant 4–vector, $a$ : form a 4–vector by the equation

$$b^\nu = a_\mu T^{\mu\nu}$$

This allows equation 3 to be re-written as

$$\frac{\partial b^\nu}{\partial x^\nu} = 0 \quad .$$

Integrating this over a hyper-volume, $V$ , of space-time gives

$$\int_V \frac{\partial b^\nu}{\partial x^\nu}\, d^4x = 0 \quad .$$

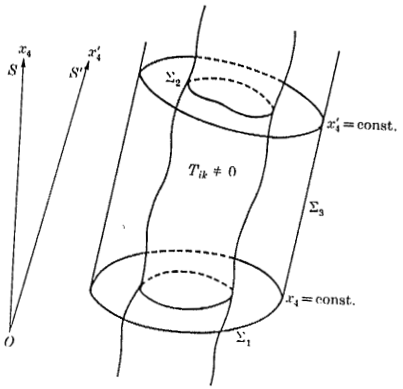
Now Gauss' theorem can be extended to apply here: if $\Sigma$ is the surface of the volume $V$ then

$$\int_\Sigma b^\nu \, d\Sigma_\nu = 0 \quad . \tag{2.4}$$

( $d\Sigma_\nu$ is the surface element 'normal' to the $x^\nu$–axis; using the anti-symmetric permutation tensor this can be formulated as

$$d\Sigma_\nu = \epsilon_{\nu\mu\rho\sigma}\, dx^\mu\, dx^\rho\, dx^\sigma \quad .)$$

The contravariance of $P$ is now a matter of choosing an appropriate volume as $V$ . The surface $\Sigma$ is taken in three pieces: $\Sigma^1$ and $\Sigma^2$ are parts of hyperplanes of constant $x^0$ and $x'^0$, respectively; $\Sigma^3$ is a surface enclosing the support of $T$ but at no point on $\Sigma^3$ is $T$ non-zero. $\Sigma^1$ and $\Sigma^2$ are chosen so that $\Sigma^2$ lies in the future of $\Sigma^1$ within $\Sigma^3$ .




The volume of integration and its boundary

Taken from Chr. Møller, *The Theory of Relativity*, O. U. P. 1972: where $x_4$ is the time-like coordinate I denote by $x^0$.

This means that equation 4 becomes

$$\int_{\Sigma^1} b^\nu \, d\Sigma_\nu - \int_{\Sigma^2} b^\nu \, d\Sigma_\nu = 0$$

or, in terms of the two reference frames,

$$\int b^0 \, dx^1\, dx^2\, dx^3 = \int b'^0 \, dx'^1\, dx'^2\, dx'^3$$

i.e.,
$$\int a_\nu T^{\nu 0}\, d^3\mathbf{x} = \int a'_\nu T'^{\nu 0}\, d^3\mathbf{x}'$$

i.e.,
$$a_\nu P^\nu = a'_\nu P'^\nu$$

– an invariant formula, whence the contravariance of $P$ . These covariant quantities are the basis of early attempts to define a classical centre of a distributed system.

In 1948 Maurice Pryce attempted an encyclopaedic enumeration of the candidate centres of position. Of the six definitions produced, only three found even partial favour with Pryce: his definitions (c), (d), and (e).

§§2.1 *Definition (c)*

> 'The co-ordinates of the mass-centre in a particular frame of reference is defined as the mean of the co-ordinates of the several particles weighted [by] their dynamical masses (energies).' [p. 63]

Or, using $T$ and the previous definitions of the total momentum, $P$ , and total angular momentum, $M$ :

$$P^0 q^\nu = \int_{\mathbb{R}^3} x^\nu T^{00}\, d^3\mathbf{x} \tag{2.5}$$

Eliminating $T$ from this expression gives

$$\begin{aligned} q^\nu &= \frac{x^0 P^\nu + M^{\nu 0}}{P^0} \\ &= \frac{q^0 P^\nu + M^{\nu 0}}{P^0} \end{aligned} \tag{2.6}$$

A formula specific to a particular frame of reference – Pryce remarks that this frame-dependency or non-covariance does not endear the definition to him. To see this, consider equation 6 applied in a frame where the coordinates are denoted by barred quantities. Then a general Lorentz transformation of the '4-vector' $\bar{q}$ will give

$$\begin{aligned} q^\nu = \Lambda^\nu_\mu \bar{q}^\mu &= \Lambda^\nu_\mu \frac{\bar{q}^0 \bar{P}^\mu + \bar{M}^{\mu 0}}{\bar{P}^0} \\ &= \Lambda^\nu_\mu \left( \frac{\bar{q}^0 \Lambda^\mu_\sigma P^\sigma + \Lambda^\mu_\rho \Lambda^0_\sigma M^{\rho\sigma}}{\Lambda^0_\sigma P^\sigma} \right) \end{aligned}$$

(But $\Lambda^\mu_\nu \Lambda^\nu_\sigma = \delta^\mu_\sigma$ )

$$= \frac{\bar{q}^0 P^\nu + \Lambda^0_\sigma M^{\nu\sigma}}{\Lambda^0_\sigma P^\sigma}$$

Writing $\bar{q}^0$ as a function of $q^0$,

$$q^0 = \frac{\bar{q}^0 P^0 + \Lambda^0_\sigma M^{0\sigma}}{\Lambda^0_\sigma P^\sigma}$$

i.e., $$\bar{q}^0 = \frac{\Lambda^0_\sigma P^\sigma q^0 - \Lambda^0_\sigma M^{0\sigma}}{P^0} \quad .$$

Whence, eliminating $\bar{q}^0$ to give an expression entirely in terms of the unbarred quantities

$$q^\nu = \frac{P^\nu}{P^0} \left( q^0 - \frac{\Lambda^0_\sigma M^{0\sigma}}{\Lambda^0_\sigma P^\sigma} \right) + \frac{\Lambda^0_\sigma M^{\nu\sigma}}{\Lambda^0_\sigma P^\sigma}$$

$$q^\nu = \frac{P^\nu}{P^0} q^0 + \frac{\Lambda^0_\sigma \left( M^{\nu\sigma} P^0 - M^{0\sigma} P^\nu \right)}{P^0 \Lambda^0_\sigma P^\sigma} \tag{2.7}$$

– a formula that directly exhibits the non-covariance of $q$ .

§§2.2 *Definition (d)*

> 'Definition (c) is first applied in a frame in which the total momentum is zero (and hence the mass-centre is at rest), and the result is then [transformed] to the [desired frame] by a Lorentz transformation.' [p. 63]

Or, starting from equation 6, with $\bar{P} = 0$ , and $\bar{P}^0 = mc$ (a frame such as this will exist for any 'physical' system):

$$P_\nu = \Lambda^\mu_\nu \bar{P}_\mu = \Lambda^0_\nu mc \quad .$$

in other words, all occurrences of $\Lambda^0_\nu$ in equation 7 can be replaced by $P_\nu / mc$ :

$$X^\nu = \frac{P^\nu}{P^0} X^0 + \frac{P_\sigma \left( M^{\nu\sigma} P^0 - M^{0\sigma} P^\nu \right)}{P^0 P_\sigma P^\sigma}$$

i.e.,

$$X^\nu = \frac{P^\nu X^0 m^2 c^2 + M^{\nu\sigma} P^0 P_\sigma - M^{0\sigma} P^\nu P_\sigma}{m^2 c^2 P^0} \tag{2.8}$$

– a formula that works for all inertial frames. Pryce notes of $X$ that, 'in spite of its appearance, [it] is relativistically covariant.' [p. 65]

§§2.3 *Definition (e)*

Pryce next assumes that the quantities $M$ and $P$ form a Lie algebra in which the Lie products are of the canonical form for the Poincaré group:

$$[P_\mu, P_\nu] = 0$$
$$[P_\sigma, M_{\mu\nu}] = P_\mu g_{\nu\sigma} - P_\nu g_{\mu\sigma}$$
$$[M_{\nu\mu}, M_{\gamma\eta}] = M_{\nu\gamma} g_{\mu\eta} + M_{\mu\eta} g_{\nu\gamma} - M_{\nu\eta} g_{\mu\gamma} - M_{\mu\gamma} g_{\nu\eta}$$

Initially, the Lie product is called a Poisson bracket, though it is significant that no definition is hazarded for such a bracket. On the basis of formulae 7 and 8 and this Lie algebra, Pryce deduces the Lie product of $q$ and $X$ with the canonical generators:

$$[q^\mu, q^\nu] = \frac{-1}{(P^0)^2} \epsilon_{\mu\nu\tau} S^\tau$$

( $\mathbf{S} = \mathbf{M} - \mathbf{q} \wedge \mathbf{P}$ )

$$[X^\mu, X^\nu] = \frac{1}{m^2} \epsilon_{\mu\nu\tau} S^\tau \tag{2.9}$$

– neither of which is the Lie product of the components of a position vector. It is a simple consequence of the last two equations that the four quantities

$$\tilde{q}^\nu = \frac{P^0 q^\nu + m X^\nu}{P^0 + m} \tag{2.10}$$

do have this property of a position vector:

$$[\tilde{q}^\nu, \tilde{q}^\mu] = 0 \quad .$$

The four quantities, $\{ \tilde{q}^\nu \}$ , form Pryce's definition (e) of the 'centre of mass'. These four quantities do not, it is admitted, form a 4–vector.

On the basis of equations 7, 8, and 10, Pryce found the following realisations of his position operators in terms of Dirac's theory of the electron:

$$\dot{\mathbf{q}} = \mathbf{x} + \frac{\hbar}{2{p_0}^2} (\mathbf{p} \wedge \sigma + im\beta\boldsymbol{\alpha})$$
$$\dot{\mathbf{X}} = \mathbf{x} + \frac{\hbar}{2m} i\beta \left( \boldsymbol{\alpha} - \frac{(\boldsymbol{\alpha} \cdot \mathbf{p})\mathbf{p}}{{p_0}^2} \right)$$
$$\tilde{\mathbf{q}} = \mathbf{x} + \frac{\hbar}{2} \left( \frac{i\beta\boldsymbol{\alpha}}{p_0} + \frac{\mathbf{p} \wedge \sigma}{p_0 (p_0 + m)} - \frac{i\beta(\boldsymbol{\alpha} \cdot \mathbf{p})\mathbf{p}}{{p_0}^2 (p_0 + m)} \right) \tag{2.11}$$

This requires a definition of $M$ ( $p$ being given) which is accomplished by the formulae

$$M^{ij} = \pm \frac{1}{2} (x^i \hat{p}_j - \hat{p}_j x^i)$$
$$M^{k0} = \frac{1}{2} (x^i \hat{p}_0 + \hat{p}_0 x^i)$$

( $M$ remains an anti-symmetric tensor by taking $M^{k0}$ , $M^{23}$ , $M^{31}$ , and $M^{12}$ positive and the rest negative.)

No attempt is made to further legitimise any of the 'observables' produced; there is no reference to any self-adjoint property, for example. Pryce concludes that the job of defining a position observables is futile, because

> 'except for particles of spin 0, it does not seem to be possible to find a definition which is relativistically covariant [ $\mathbf{X}$ is; $\tilde{\mathbf{q}}$ is not] and at the same time yields commuting co-ordinates [ $\mathbf{X}$ does not; $\tilde{\mathbf{q}}$ does].' [p. 69]

Pryce favoured the use of $\mathbf{X}$ over $\tilde{\mathbf{q}}$ for its covariance; and he further claims that for spin-0 the components of $\mathbf{X}$ commute, though this is doubly dubious for: Newton and Wigner's operators do not behave this way, and, secondly, in equation 9 the commutator of two components of $\mathbf{X}$ is a component of the internal angular momentum, $S$ –an entirely classical quantity – not spin.

## §3 The Newton–Wigner Position Operator Re-derived

T. D. Newton, E. P. Wigner, *Localized states for elementary systems* Rev. Mod. Phys. **21** (1949) 400–406.

In 1949 Newton and Wigner derived the position operators that now bear their names. Any assessment of this work must, of course, take account of the considerable developments made in the intervening years. Indeed, it is arguable – an argument that I shall accept here – that the approach was not as rigorous as it could have been even in 1949. Von Neumann's *Mathematical Foundations of Quantum Mechanics* had been published in 1932, in which the unphysical 'eigen-functions' for observables with continuous spectra were effectively excised. Yet here, some 17 years later, Newton and Wigner casually discuss (my emphasis), 'the *state* (or *states*) ... for which the three space coordinates are zero at $t = 0$.' This reaches its most sublimely ludicrous statement in the authors' equation 7:

$$\text{`} \quad \psi^2 = (2\pi)^{-3} p_0 \quad \text{'}$$

where the factor of $(2\pi)^{-3}$ can have no conceivable significance since '(a)s was anticipated, $(\psi, \psi)$ is *infinite*'. Such quibbles aside, the merit of the approach was to show explicitly that the symmetries demanded were present. The major failing is that the number of symmetries required (isotropy and homogeneity of space) is less than the number of interest (Lorentz boosts, homogeneity of time). The formulation that appears here is less defective in discussing states and observables, but is no less so in dealing with symmetries.

Following Wigner's 1939 paper,[4] initially the whole theory is cast in the spectral representation space of the 4–momentum, $(p_\mu)$ , on which an irreducible unitary representation of the orthochronous Poincaré group is defined:

$$\tilde{\mathcal{H}} = \tilde{\mathcal{H}}(0, m) = L^2\left(\Re^3, \frac{d^3\mathbf{p}}{2\sqrt{|\mathbf{p}|^2 + m^2c^2}}\right) \quad ,$$

for spin–0 particles of rest-mass $m$ . Analogous considerations apply for non-zero spin, so for ease of presentation only the spin–0 theory will be dealt with in the remainder of this section.

The crux of the problem might be said to be the fact that Position is not one of the generators of the Poincaré group. The next step is, therefore, to find a suitable candidate for the rôle of Position Observable. There are a variety of ways of going about this.

In 1949 Newton and Wigner, effectively, sought the projector–valued spectral function for position, $E(\hat{\mathbf{x}}; \mathbf{a})$ . To do this, they deduced the 'eigen-function' of position corresponding to a particle at the origin of coordinates (giving them $E(\hat{\mathbf{x}}; 0)$ ) which can thence be used to define the rest of the spectral function by translations:

$$E(\hat{\mathbf{x}}; \mathbf{a}) = \exp\left(\frac{i}{\hbar}\hat{\mathbf{p}} \cdot \mathbf{a}\right) E(\hat{\mathbf{x}}; 0) \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}} \cdot \mathbf{a}\right) \quad . \tag{3.1}$$

By contrast, Schweber[5][6], found a much shorter derivation by considering the operator which is the 3–momentum representation of position in the Galilei-invariant theory:

$$i\hbar\nabla_p \quad ,$$

and thence finding the correction term to make it symmetric in $\tilde{\mathcal{H}}$.

Whilst both methods give the same momentum space representation for $\hat{x}$ , they are not, in my opinion, as convincing as they could be. A more careful line of thought might be to consider at least some of the transformation properties of the putative observable, $\hat{x}$ . Take the translations of $\hat{x}$ :

$$\exp\left(\frac{i}{\hbar}\hat{\mathbf{p}} \cdot \mathbf{a}\right) \hat{\mathbf{x}} \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}} \cdot \mathbf{a}\right) = \hat{\mathbf{x}} + \mathbf{a} \quad , \tag{3.2}$$

or, equivalently,

$$[\hat{x}_j, \hat{p}_k] = i\hbar\delta_{jk} \quad . \tag{3.3}$$

It is known that, for $\hat{\mathbf{p}}$ a multiplication operator, the most general solution is

$$\hat{x}_j = i\hbar\frac{\partial}{\partial p^j} + f_j(\mathbf{p}) \quad . \tag{3.4}$$

A considerable constraint upon the functions $\{ f_j \}$ is the requirement that $\hat{x}$ be a symmetric operator in $\tilde{\mathcal{H}}$ . Thus it is demanded that

$$(\tilde{\psi}, \hat{x}\tilde{\phi})_{\tilde{\mathcal{H}}} = (\hat{x}\tilde{\psi}, \tilde{\phi})_{\tilde{\mathcal{H}}} \quad , \tag{3.5}$$

i.e.,

$$\int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \tilde{\Psi}^*(p_{0+},\mathbf{p}) \left( i\hbar \frac{\partial}{\partial p^j} + f_j(\mathbf{p}) \right) \tilde{\Phi}(p_{0+},\mathbf{p})$$

$$= \int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \left[ \left( i\hbar \frac{\partial}{\partial p^j} + f_j(\mathbf{p}) \right) \tilde{\Psi} \right]^* \tilde{\Phi}$$

$$= \int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \left( -i\hbar \frac{\partial \tilde{\Psi}^*}{\partial p^j} + \left( f_j(\mathbf{p}) \tilde{\Psi} \right)^* \right) \tilde{\Phi}$$

integrating by parts

$$= \int_{\Re^3} d^3\mathbf{p}\, \tilde{\Psi}^* i\hbar \frac{\partial}{\partial p^j} \left( \frac{\tilde{\Phi}(\mathbf{p})}{2p_{0+}} \right) + \int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \left( f_j(\mathbf{p}) \tilde{\Psi} \right)^* \tilde{\Phi}$$

Now

$$\frac{\partial}{\partial p^j} \left( \frac{1}{2p_{0+}} \right) = \frac{-1}{4\left(|\mathbf{p}|^2 + m^2c^2\right)^{3/2}} \cdot 2p_j \quad . \tag{3.6}$$

So,

$$\int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \tilde{\Psi}^* \left( i\hbar \frac{\partial}{\partial p^j} + f_j(\mathbf{p}) \right) \tilde{\Phi} = \int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \tilde{\Psi}^* \left( i\hbar \frac{\partial \tilde{\Phi}}{\partial p^j} - \frac{i\hbar p_j \tilde{\Phi}}{|\mathbf{p}|^2 + m^2c^2} \right) +$$

$$\int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \left( f_j(\mathbf{p}) \tilde{\Psi} \right)^* \tilde{\Phi}$$

$$= \int_{\Re^3} \frac{d^3\mathbf{p}}{2p_{0+}} \tilde{\Psi}^* \left( i\hbar \frac{\partial}{\partial p^j} + f_j^* - \frac{i\hbar p_j}{|\mathbf{p}|^2 + m^2c^2} \right) \tilde{\Phi}$$

Clearly, if $\hat{x}$ is symmetric for all states $\tilde{\Phi}$ and $\tilde{\Psi}$ , then, except on a set of measure zero,

$$f_j = f_j^* - \frac{i\hbar p_j}{|\mathbf{p}|^2 + m^2c^2} \quad ; \tag{3.7}$$

$\Rightarrow$

$$f_j - f_j^* = 2i\Im(f_j) = -\frac{i\hbar p_j}{|\mathbf{p}|^2 + m^2c^2} \quad . \tag{3.8}$$

Thus, it has been shown that, if $\mathbf{h} : \Re^3 \to \Re^3$ ,

$$\mathbf{f}(\mathbf{p}) = \mathbf{h}(\mathbf{p}) - \frac{i\hbar \mathbf{p}}{2\left(|\mathbf{p}|^2 + m^2c^2\right)} \quad . \tag{3.9}$$

There will be additional constraints, i.e., restrictions on $\mathbf{h}$ , e.g., if it is required that $\hat{x}$ be self-adjoint. One obvious candidate is obtained by putting

$$\mathbf{h} = \mathbf{0} \quad . \tag{3.10}$$

This gives the Newton–Wigner operator for zero spin:

$$\hat{\mathbf{x}} = i\hbar \nabla_p - \frac{i\hbar \mathbf{p}}{2\left(|\mathbf{p}|^2 + m^2c^2\right)} \quad . \tag{3.11}$$

In fact, regardless of the form of $\mathbf{h}$ , the 'eigen-vectors' of the position operator found here are the functions

$$\varsigma_x(\mathbf{p}) = \sqrt{p_{0+}} \exp\left( \frac{i}{\hbar} x^\mu p_\mu \right) \quad , \tag{3.12}$$

which have 'eigen-values' given by the equation

$$\hat{\mathbf{x}}\zeta_x = \left(\mathbf{x} - \frac{x^0\mathbf{p}}{p_0}\right)\zeta_x \quad . \tag{3.13}$$

The second term in the eigen-value arises from the evolution of position in time.

The transformation to configuration space, that is to the spectral representation space of $\hat{\mathbf{x}}$ , is then accomplished, in the usual way, by the formula:

$$\begin{aligned}\psi(x) &= \langle x|\psi\rangle \\ &= \frac{1}{N}\int_{\Re^3}\frac{d^3\mathbf{p}}{2p_{0+}}\zeta_x^*(\mathbf{r})\tilde{\Psi}(p_{0+},\mathbf{p}) \\ &= \frac{1}{N}\int_{\Re^3}\frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}}\exp\left(\frac{i}{\hbar}x^\mu p_\mu\right)\tilde{\Psi}(p_{0+},\mathbf{p})\end{aligned}$$

It is clear from the dependence on $x$ in this formula that all configuration space wave-functions defined in this way will be solutions of the Klein–Gordon equation, in at least the weak sense.

To make this coordinate representation, or 'configuration space', into a Hilbert space, $\mathcal{H}$ , this transformation is used to define the inner-product of configuration space wave-functions, by way of their momentum space counter-parts. Thus

$$(\psi(x),\phi(x))_{\mathcal{H}} = (\tilde{\psi}(\mathbf{p}),\tilde{\phi}(\mathbf{p}))_{\tilde{\mathcal{H}}}$$

Now

$$\begin{aligned}(\tilde{\psi}(\mathbf{p}),\tilde{\phi}(\mathbf{p}))_{\tilde{\mathcal{H}}} &= \int_{\Re^3}\frac{d^3\mathbf{p}}{2p_{0+}}\tilde{\Psi}^*(p_{0+},\mathbf{p})\tilde{\Phi}(p_{0+},\mathbf{p}) \\ &= 2\int_{\Re^3}\frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}}\int_{\Re^3}\frac{d^3\mathbf{p}'}{2\sqrt{p'_{0+}}}\delta(\mathbf{p}-\mathbf{p}')\tilde{\Psi}^*(p'_{0+},\mathbf{p}')\tilde{\Phi}(p_{0+},\mathbf{p})\end{aligned}$$

But

$$\delta(\mathbf{p}-\mathbf{p}') = \frac{1}{(2\pi\hbar)^3}\int_{\Re^3}d^3\mathbf{x}\,\exp\left(\frac{i}{\hbar}\mathbf{x}\cdot(\mathbf{p}-\mathbf{p}')\right) \tag{3.14}$$

So

$$\begin{aligned}&(\tilde{\psi}(\mathbf{p}),\tilde{\phi}(\mathbf{p}))_{\tilde{\mathcal{H}}} = \\ &= 2\int_{\Re^3}\frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}}\int_{\Re^3}\frac{d^3\mathbf{p}'}{2\sqrt{p'_{0+}}}\frac{1}{(2\pi\hbar)^3}\int_{\Re^3}d^3\mathbf{x}\,\exp\left(\frac{i}{\hbar}\mathbf{x}\cdot(\mathbf{p}-\mathbf{p}')\right)\tilde{\Psi}^*(p'_{0+},\mathbf{p}')\tilde{\Phi}(p_{0+},\mathbf{p}) \\ &= \frac{2}{(2\pi\hbar)^3}\int_{\Re^3}\left(\int_{\Re^3}\frac{d^3\mathbf{p}'}{2\sqrt{p'_{0+}}}\exp\left(\frac{i}{\hbar}\mathbf{x}\cdot\mathbf{p}'\right)\tilde{\Psi}(p'_{0+},\mathbf{p}')\right)^* \times \\ &\qquad\left(\int_{\Re^3}\frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}}\exp\left(\frac{i}{\hbar}\mathbf{x}\cdot\mathbf{p}\right)\tilde{\Phi}(p_{0+},\mathbf{p})\right)d^3\mathbf{x} \qquad (3.15) \\ &= \frac{2N^2}{(2\pi\hbar)^3}\int_{\Re^3}\psi^*(\mathbf{x})\phi(\mathbf{x})\,d^3\mathbf{x} = (\psi(x),\phi(x))_{\mathcal{H}} \quad . \qquad (3.16)\end{aligned}$$

The transformation to configuration space can now be 'normalized' by setting

$$N^2 = \frac{(2\pi\hbar)^3}{2} \tag{3.17}$$

to give

$$\psi(x) = \sqrt{\frac{2}{(2\pi\hbar)^3}} \int_{\Re^3} \frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}} \exp\left(\frac{i}{\hbar} x^\mu p_\mu\right) \tilde{\Psi}(p_{0+}, \mathbf{p}) \tag{3.18}$$

and so configuration space states lie in the Hilbert space

$$\mathcal{H} = L^2(\Re^3, d^3\mathbf{x})$$

The discrepancy between equations 15 and 18 is overcome by invoking the temporal homogeneity of space-time, and therefore the possibility of always taking $x^0 = 0$ by shifting the origin of time. This can be made clearer by re-writing equation 15 as:

$$(\psi(x), \phi(x))_{\mathcal{H}} = \frac{2}{(2\pi\hbar)^3} \int_{\Re^3} \left( \int_{\Re^3} \frac{d^3\mathbf{p}'}{2\sqrt{p'_{0+}}} \exp\left(\frac{i}{\hbar} x^\mu p'_\mu\right) \tilde{\Psi}(p'_{0+}, \mathbf{p}') \right)^* \times \left( \int_{\Re^3} \frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}} \exp\left(\frac{i}{\hbar} x^\mu p_\mu\right) \tilde{\Phi}(p_{0+}, \mathbf{p}) \right) d^3\mathbf{x}$$

in which, when the $\delta$–function in $\mathbf{p} - \mathbf{p}'$ is integrated out, the exponentials involving $x^0$ cancel.

It is then only a brief step to conclude that the configuration space (i.e., spatial) probability density is

$$\begin{aligned} \rho(x) &= |\psi(x)|^2 \\ &= \int_{\Re^3} \frac{d^3\mathbf{p}}{2\sqrt{p_{0+}}} \int_{\Re^3} \frac{d^3\mathbf{p}'}{2\sqrt{p'_{0+}}} \exp\left(\frac{i}{\hbar} x^\mu (p_\mu - p'_\mu)\right) \tilde{\Psi}^*(p'_{0+}, \mathbf{p}') \tilde{\Psi}(p_{0+}, \mathbf{p}) \end{aligned} \tag{3.19}$$

Barut and Malin, however, claim that $\rho$ is not the zero or time-like component of a 4–vector;[7] this claim is easily proved. Consider the Poincaré transformation (without time reflection):

$$p_\mu = \Lambda^\nu_\mu \bar{p}_\nu \quad . \tag{3.20}$$

The invariant volume element is

$$\frac{d^3\mathbf{p}}{p_{0+}} = \frac{d^3\bar{\mathbf{p}}}{\bar{p}_{0+}}$$

so

$$\begin{aligned} \frac{d^3\mathbf{p}}{\sqrt{p_{0+}}} &= \frac{\sqrt{p_{0+}(\bar{p})}}{\bar{p}_{0+}} d^3\bar{\mathbf{p}} \\ &= \frac{\sqrt{\Lambda^\nu_0 \bar{p}_\nu}}{\bar{p}_{0+}} d^3\bar{\mathbf{p}} \\ &\neq \frac{d^3\bar{\mathbf{p}}}{\sqrt{\bar{p}_{0+}}} \quad . \end{aligned}$$

A result that is not particularly surprising.

As Newton and Wigner indicated, the problem is with Lorentz boosts, and would seem to form a conclusive case against the operator they deduced. It is hard to see where any enthusiasm could be found for the Newton–Wigner operators, though there is certainly nothing better on offer. Indeed, Wigner has even questioned to what extent 'position' can occur in a Poincaré–invariant theory at all.[8]

As a footnote on the paper by Newton and Wigner, I can find no reasonable basis for the repeated contention that the operators derived are the same as Pryce's case (c) (formulae 2.10 and 2.8):

> '...the position operators to which our postulates lead necessarily commute with each other so that only Pryce's case (c) can be used for comparison. In fact, our $q^k$ is identical with his $\tilde{q}^k$ .' [p. 403]

Since the Newton–Wigner operators correspond to multiplication by $x$ in the coordinate representation generated by those operators, I see no way of reconciling the two formulae. There is a

considerable gulf between Pryce's classically motivated and Newton & Wigner's group theoretically motivated strategies.

## §4 The Klein–Gordon Current

The Newton–Wigner operator produces a coordinate representation and probability density that is at variance with the entities that usually bear those titles in discussion of the Klein–Gordon equation. By the same method as for the Schrödinger equation, the Klein–Gordon probability 4–current (so-called) is

$$j_\nu(x) = i\hbar\psi^* \frac{\partial\psi}{\partial x^\nu} - i\hbar\psi \frac{\partial\psi^*}{\partial x^\nu} \quad . \tag{4.1}$$

which satisfies a continuity equation:

$$\Box^\nu j_\nu = 0 \quad .$$

This can also be obtained from the momentum representation assumed by Newton and Wigner; though now the transformation formula between the representations is the invariant

$$\psi(x) = \frac{1}{N} \int_{\Re^3} \frac{d^3\mathbf{p}}{p_{0+}} \exp\left(\frac{i}{\hbar} x^\mu p_\mu\right) \tilde{\Phi}(p_{0+}, \mathbf{p})$$

and the momentum representation of the position operator that goes with this is

$$i\hbar\nabla_p + \mathbf{h}(\mathbf{p})$$

which is not a symmetric operator. This failing is compounded by the fact that, even limited to positive-energy states alone, $j_0$ is not a positive-definite quantity (for a more detailed account, see §6 below). Indeed, because the wave-functions in the coordinate representation of the Newton–Wigner operator satisfy the Klein–Gordon equation, the 'current' $j$ can be obtained in direct competition with the Newton–Wigner 'probability density', $\rho$ (equation 3.19). The contest is, by no means, one-sided, with neither candidate possessing all reasonable properties.

## §5 Gordon Fleming's Redemption of the Newton–Wigner Operators

G. N. Fleming, *Covariant position operators, spin, and locality*, Phys. Rev. **137B** (1965) 185–197.

In 1965 Gordon Fleming took on the challenge of showing that Newton and Wigner were wrong about their own operators; he set about showing that, contrary to the best opinions, Newton and Wigner had produced fully covariant operators. It is my firm belief that this claim is not tenable without a dilution of the idea of Relativity to the stage when it is indistinguishable from Lorentz's aether interpretation. To explore Fleming's thesis and my counter-claim, a representative selection of quotations will be analysed.

Before beginning, it is just necessary to note a somewhat unusual short-hand that Fleming introduces: if the components of some candidate position operator commute, Fleming calls this operator *local*. To draw attention to this, rather obfuscatory, new meaning, the word 'local' will be emphasized (as *local*, *locality*, &c.).

Fleming begins by casting doubts:

> 'If the correspondence principle is emphasized and used to derive the transformation properties which position operators must have then it *seems* impossible to construct a *local* operator with those transformation properties. If, on the other

> hand, *locality* is demanded at the outset, the resulting unique operator *seems* to obey bizarre transformation properties under [Lorentz boosts].' [p. 188]

Previous authors, such as Newton and Wigner, did not mince around with 'bizarre'; the transformations were not those of covariant quantities. It is worth noting that the uniqueness of the Newton–Wigner operators relies on a 'regularity assumption', amongst the other, more straightforward, axioms.

> 'It will furthermore be shown that the differences which exist between the various operators are due to their describing the position of a particular space-time point, defined in an invariant manner, in some instances; and describing the location of a dynamical property, which property depends for its location in space-time on the frame from which it is observed, in other instances.' [p. 188]

The wording of the second type of 'position' sounds not at all sensible. Firstly, 'the location of a dynamical property' must be presumed to mean that some weighted average of coordinates is being taken – the weighting being determined by the 'dynamical property'. Further, the last phrase describing this second type, expressing the dependence on a preferred coordinate frame, smacks of an unrelativistic quantity. It is worth mentioning that any quantity (with 1, 4, 16, or $4^n$ 'components') can masquerade as a tensor by the *ad hoc* assertion of the right transformation law. Having found this second sort of 'position' to be dodgy, I can now allow Fleming to continue:

> 'The Newton–Wigner operator is [of] the second kind and the frame dependence of the point thereby *localised* is frequently said to indicate the noncovariance of the operator. Such terminology is unfortunate, however, since no legitimate covariance requirement has been violated.... What is demanded by such terminology is that the *point localised* by the operator be *invariant* and there is no *a priori* reason [to expect] a dynamical property to have an invariant location.' [p. 188–189]

– which all hinges on the precise nature of the 'dynamical property'. It is Fleming's contention that all these 'dynamical properties' are the result of integrals over constant-time hyperplanes – a hypothesis that immediately raises the problem of the distinct manner in which space and time are treated in Galilei–invariant quantum theories, and which Fleming is determined to perpetuate in a Poincaré–invariant theory:

> 'The fact that the time variable is a $c$-number throughout quantum theory makes a relativistic treatment of position measurements awkward. The source of the $c$-number character of the time coordinate in relativistic quantum theory is the preoccupation with *instantaneous* ( $t = \text{const.}$ ) hyperplanes in the discussion of position measurements.' [p. 189]

The 'preoccupation' is entirely borrowed from the Schrödinger–Heisenberg theory, and not at all a necessary feature for a Poincaré–invariant theory.

Just as a number of other authors have, Fleming chose a new invariant for his evolution parameter, $\tau$ . To denote the hyperplane on which measurements are to be conducted, a 4–vector, $\eta$ , orthogonal to this hyperplane is found:

$$\eta^\nu \eta_\nu = 1 \quad . \tag{5.1}$$

A 4–vector position observable, $\chi$ , is then obliged to satisfy the constraint

$$\chi_\nu \eta^\nu = \tau \quad , \tag{5.2}$$

and the transformation law

$$\langle \phi' | \hat{\chi}_\mu(\eta', \tau') | \phi' \rangle = a^\nu_\mu \langle \phi | \hat{\chi}_\nu(\eta, \tau) | \phi \rangle + a_\mu$$

( $\eta'_\mu = a^\nu_\mu \eta_\nu$ , $\tau' = \tau + \eta'^\mu a_\mu$ .)

> '...[this equation] relates position coordinates on the *same* hyperplane as seen in different frames.... This last point raises the question of how the manifestly covariant operators [*sic*] depend on the parameters $\eta$ and $\tau$ .' [p. 190]

This certainly needs scrutinising when Pryce, whom Fleming cites and whose notation Fleming is using, was quite clear that a dependence on $\eta$ was the sign of faulty symmetry properties.

If any value is completely tied to a single, specific, constant-time hyperplane, it is reasonable to describe this as frame dependent even if it is expressed in terms of an arbitrary coordinate system.

Indeed, the use of quantities dependent on a single coordinate frame is an unhappy mix of the notation of Relativity and the philosophy of a preferred frame of reference. Fleming tries to escape from having to justify preferring one specific frame by allowing sets of quantities to be defined in all inertial frames; quantities unrelated by any transformation law. The idea of an invariant formulation of the laws of physics is supplanted by the notion that because a set of definitions can be implemented in any frame then that will somehow be adequate.

Consider now the presentation of equation 2 by Fleming. He describes $\tau$ as an invariant in the text but admits in a footnote that this is not so. Starting from the supposition that $\chi$ is a position 4–vector and $\eta$ a difference of such vectors, it is readily apparent that $\chi_\nu \eta^\nu$ is not an invariant. Under a general Poincaré transformation, $\{\Lambda, \bar{a}\}$ , this means

$$\chi_\nu = \Lambda^\mu_\nu \overline{\chi}_\mu + \bar{a}_\nu$$
$$\eta^\nu = \Lambda^\nu_\mu \overline{\eta}^\mu$$

So

$$\chi_\nu \eta^\nu = \left(\Lambda^\mu_\nu \overline{\chi}_\mu + \bar{a}_\nu\right)\left(\Lambda^\nu_\mu \overline{\eta}^\mu\right)$$
$$= \overline{\chi}_\nu \overline{\eta}^\nu + \bar{a}_\nu \Lambda^\nu_\mu \overline{\eta}^\mu$$

This means that, starting from equation 2 as the constraint defining a set of constant-time hyperplanes, an equivalent description of the same hyperplane, now in terms of the barred coordinate system, is the equation

$$\tau = \overline{\chi}_\nu \overline{\eta}^\nu + \bar{a}_\nu \Lambda^\nu_\mu \overline{\eta}^\mu$$

or

$$\overline{\chi}_\nu \overline{\eta}^\nu = \tau - \bar{a}_\nu \Lambda^\nu_\mu \overline{\eta}^\mu$$

– the right-hand side of which is a constant that changes as one hyperplane is supplanted by the next, as $\tau$ is varied, at the same rate in all frames. The quantity $\tau$ remains the time-like coordinate in the frame of definition – the frame in which $\eta = (1, 0)$ .

Parenthetically, it should be noticed that Fleming's 'Fig. 1' and 'Fig. 2' are incorrect in their portrayal of $\eta'$ , which is orthogonal to its hyperplane with respect to the Minkowski metric, and so, in the coordinate system illustrated, will not be perpendicular to its hyperplane.

Having established the true nature of $\tau$ , the next quotation becomes easier to grasp.

> 'In the conventional formalism one demands
>
> $$\frac{d}{dt}\langle\phi|\chi(t)|\phi\rangle = \langle\phi|\mathbb{P}/P_0|\phi\rangle c \qquad (5.3)$$
>
> where $P_\nu$ is the total four–momentum vector. The manifestly covariant generalisation [*sic*] of this result is

$$\frac{\partial}{\partial\tau}\langle\phi|\overline{\chi}_\mu(\overline{\eta},\tau)|\phi\rangle = \langle\phi|\bar{P}_\mu/\overline{\eta}^\nu\bar{P}_\nu|\phi\rangle \qquad (5.4)$$

' [p. 190]

There is no clue as to which 'conventional formalism' is being employed: there is no widely agreed Poincaré–invariant quantum mechanics to be conventional; the usual quantum mechanics of Schrödinger and Heisenberg would not have the factor of $c$ and $P_0$ would be the mass, $m$ . This leaves the conventional, classical mechanics of point particles from which, by analogy (the 'correspondence principle'), one might demand equation 3.

The 'covariant generalization' now being invoked is neither covariant, nor are covariant formulae in any typical need of generalization. On the left-hand side it is obviously permissible to find the derivative with respect to a certain time coordinate regardless of the coordinate system in use, so there is no change here. On the right-hand side the transformation equation

$$P_\nu = \Lambda^\mu_\nu \bar{P}_\mu$$

has been utilised; since

$$(\eta) = (1,\mathbf{0})$$

so, it can be deduced,

$$\overline{\eta}^\nu = \Lambda^\nu_0$$

and so

$$P_0 = \Lambda^\nu_0 \bar{P}_\nu = \overline{\eta}^\nu \bar{P}_\mu$$

whence

$$\frac{P_\nu}{P_0} \quad \text{'transforms to'} \quad \frac{\bar{P}_\mu}{\overline{\eta}^\nu \bar{P}_\mu} \ .$$

This is exactly as Pryce proceeded in his case (c). Equation 4 is, therefore, the expression of equation 3 in some other inertial frame, as opposed to the version of equation 3 for this particular frame.

The best statement of what Fleming believes to be going on is the introduction to his third section:

> 'Consider a position four–vector $\chi_\mu(\eta,\tau)$ in the classical limit [or, even, classically]. For a fixed value of $\eta$ and variable $\tau$ this four–vector traces out a world-line in space-time. In general, changing $\eta$ will alter the world-line, i.e., the location of the world-line will depend on the orientation of the space-like hyperplane on which the points of the world-line are observed. Of course, one would never expect such behaviour of a four–vector describing the position of a "point" particle, but it is quite reasonable and, in fact, the case that extended systems possess *localisable* dynamical properties which depend, for their location, on the orientation of the hyperplanes on which they are observed. Nevertheless ... those four–vectors which describe the motion of points in the system which have been defined in an invariant manner [are] of interest.... [Such a four–vector will] be called a "point" four–vector.' [p. 191]

I feel the generous admission that invariant quantities are 'nevertheless of interest' will come as a great relief to serious students of relativity. The remainder of the paragraph, however, springs from a blind acceptance of the sweeping assumption, stated twice, that physical quantities are measured on constant-time hyperplanes. This is not an uncommon assumption, arising as it does from the familiar and well-established Galilei–invariant physics. The crucial difference is that constant-time hyperplanes are not invariant under Poincaré transformations. Consequently, a constant-time hyperplane is not a set of points with any particular significance: it is merely a set with a simple defining property, in its frame of definition.

The claim that anything is determined by a measurement on a constant-time surface is patently ridiculous. Now, it is true that the total 4–momentum and total angular momentum tensor of a classical field are *computed* by integration over a constant-time hyperplane (as shown in §2, above); but these quantities do not depend on this surface: the same tensors may be obtained from any constant-time surface in any inertial frame, if not from more general hypersurfaces. It should also be noted that a computation from a model in no way implies any 'observation' or 'measurement'. The

aspect of the geometry and physics being exploited is that a constant-time hyperplane consists of a complete set of entirely independent events. Completeness here means that it is assumed the physical system is determined by the values of the various fields on this set of events. Such completeness cannot always be assumed.

To conclude, Fleming has tried to distinguish between 'point' operators and operators giving the location of some 'dynamical property' without having to call either non-invariant. His argument is not credible. By his own admission the Newton–Wigner operators are not 'point' operators. I conclude that this means they are not invariant, just as Newton and Wigner had said.

## §6 The 4–Currents of Gromes, Gerlach & Petzold

B. Gerlach, D. Gromes, J. Petzold, *The construction of definite expressions for the particle density of the Klein–Gordon field*, Z. Phys. **204** (1967) 1–11. (For a translation, see Appendix B.)

There are several ways in which the notion of locality can be said to enter the Galilei–invariant quantum theory. One of these is by the probability current, $\mathbf{j}$ , and probability density, $\rho$ , which obey a conservation equation:

$$\frac{\partial \rho}{\partial t} + \nabla.\mathbf{j} = 0 \tag{6.1}$$

In considering a Poincaré–invariant quantum mechanics, one obvious analogy to exploit is this equation. It requires but a small effort to realise that the spatial symmetry of the Galilean $\mathbf{j}$ can lead to an object that conforms to the whole Poincaré group if the four-component object:

$$j^\mu = (c\rho, \mathbf{j}) \quad , \tag{6.2}$$

is a contravariant tensor of the first rank. Indeed, one might feel compelled to this conclusion by the fact that the continuity equation can be written, with this assumption about $j^\mu$ , as

$$\frac{\partial j^\mu}{\partial x^\mu} = 0 \quad . \tag{6.3}$$

To produce the quantum mechanical version of this, one again proceeds by analogy: this time following the method that gives the quantum probability current and density in the Schrödinger–Born theory. Starting from the Klein–Gordon equation:

$$\hbar^2 \frac{\partial^2 \psi}{\partial (x^0)^2} - \hbar^2 \nabla^2 \psi + m^2 c^2 \psi = 0 \tag{6.4}$$

then

$$\begin{aligned} 0 &= \psi^* \times (\text{eq. } 4) - \psi \times (\text{eq. } 4)^* \\ &= \hbar^2 \left( \psi^* \frac{\partial^2 \psi}{\partial (x^0)^2} - \psi \frac{\partial^2 \psi^*}{\partial (x^0)^2} \right) - \hbar^2 \left( \psi^* \nabla^2 \psi - \psi \nabla^2 \psi^* \right) \\ &= \hbar^2 \left( \frac{\partial}{\partial x^0} \left( \psi^* \frac{\partial \psi}{\partial x^0} - \psi \frac{\partial \psi^*}{\partial x^0} \right) - \nabla . \left( \psi^* \nabla \psi - \psi \nabla \psi^* \right) \right) \quad . \end{aligned} \tag{6.5}$$

Thus, writing

$$\begin{aligned} c\rho &= \psi^* \frac{\partial \psi}{\partial x^0} - \psi \frac{\partial \psi^*}{\partial x^0} \quad , \\ \mathbf{j} &= -(\psi^* \nabla \psi - \psi \nabla \psi^*) \quad , \end{aligned} \tag{6.6}$$

one gets equations 1 and 3.

In fact, any constant multiple of $(c\rho, \mathbf{j})$ can be taken to be the 4–current. Conventionally, the Klein–Gordon current is

$$s^\mu = \frac{i\hbar}{2} (c\rho, \mathbf{j}) \quad . \tag{6.7}$$

One drawback of $s^\mu$ is that $s^0$ –which one might hope to be a probability density – is not positive-definite. The justification for taking equation 7 is that one can also write the 4–current as

$$s^\mu = s^\mu_+ + s^\mu_- \quad , \tag{6.8}$$

where

$$s^\mu_\pm = \frac{1}{(2\pi\hbar)^3}\int_{\mathfrak{Re}} \frac{d^3\mathbf{p}}{p_{0\pm}}\frac{d^3\mathbf{p}'}{p'_{0\pm}}\tilde{\psi}^*_\pm(\mathbf{p})\tilde{\psi}_\pm(\mathbf{p}')\exp\left(\frac{i(p_\mu - p'_\mu)x^\mu}{\hbar}\right)\frac{(p^\mu + p'^\mu)}{2} \quad , \tag{6.9}$$

and where

$$p_{0\pm} = \pm\sqrt{|\mathbf{p}|^2 + m^2c^2} \quad . \tag{6.10}$$

Whilst it might be thought that the presence of $p_{0-}$ is the cause of the negative values of $s^0$ , Blokhintsev[9] and Gerlach, Gromes & Petzold have shown that. in fact, $s^0_+$ is not positive-definite.[10] This is at odds with the positive expression Schweber obtained for $s^0$ :[6]

$$s^0 = \frac{E}{mc^2}\phi^*\phi \quad , \tag{6.11}$$

though to get this it must be assumed that $\phi$ is a mono-energetic state, so avoiding the use of the energy operator, $\hat{E}$ . In general, $s^0$ is a Lebesgue–Stieltjes integral over the energy spectrum. Considering the next most simple case, let $\phi$ be a 'state' involving only two energies:

$$\begin{aligned} \phi &= \phi_1 + \phi_2 \\ \hat{E}\phi_i &= e_i\phi_i \end{aligned} \tag{6.12}$$

Then

$$\begin{aligned} \phi^*\hat{E}\phi &= (\phi_1^* + \phi_2^*)(e_1\phi_1 + e_2\phi_2) \\ &= e_1\phi_1\phi_1^* + e_2\phi_2\phi_2^* + e_1\phi_1\phi_2^* + e_2\phi_2\phi_1^* \end{aligned}$$

Now if

$$\begin{aligned} \phi_1 &= a + ib \\ \phi_2 &= c + id \end{aligned}$$

then

$$\begin{aligned} \phi^*\hat{E}\phi &= e_1(a^2+b^2) + e_2(c^2+d^2) + e_1(ac + bd + i(bc - da)) + e_2(ac + bd - i(bc - da)) \\ &= a^2e_1 + ac(e_1+e_2) + c^2e_2 + b^2e_1 + bd(e_1+e_2) + d^2e_2 + i(bc - da)(e_1 - e_2) \end{aligned}$$

The difference between $\phi^*\hat{E}\phi$ and $\phi\hat{E}\phi^*$ is that the last term appears negated. Thus

$$s^0 \propto a^2e_1 + ac(e_1+e_2) + c^2e_2 + b^2e_1 + bd(e_1+e_2) + d^2e_2 \quad . \tag{6.13}$$

So, if $e_2 > e_1$ , i.e., $e_2 = (1+k)e_1$ where $k > 0$ .

$$s^0 \propto e_1(a+c)^2 + e_1(b+d)^2 + e_1k(ac + bd + c^2 + d^2) \quad . \tag{6.14}$$

There is no reason that $ac + bd$ cannot be negative, nor, therefore, that $s^0$ may not be negative locally†.

Now it would obviously be much easier if there was a positive-definite probability density. Indeed, if there is some 4–current, $\tilde{s}^\mu$ , such that $\tilde{s}^0$ is positive-definite in every coordinate frame‡, then Gerlach, Gromes & Petzold have shown that it satisfies a causality condition:[11]

$$\int_{V-\tau}\tilde{s}^0(x^0 \pm r, \mathfrak{x})\,d^3x \le \int_V \tilde{s}^0(x^0, \mathfrak{x})\,d^3x \le \int_{V+\tau}\tilde{s}^0(x^0 \pm r, \mathfrak{x})\,d^3x \quad . \tag{6.15}$$

$$\begin{aligned} V + \tau &= \{\, x^\mu : (\exists y^\mu \in V) \quad |\mathbf{x} - \mathbf{y}| \le \tau \,\} \\ V - \tau &= \{\, x^\mu \in V : (\forall y^\mu \in \partial V) \quad |\mathbf{x} - \mathbf{y}| \ge \tau \,\} \quad . \end{aligned}$$

† Try $e_1 = 10^{-6}$, $c = d = 0.1$, $k = 1$ and $a = b = -10$.

‡ Alternatively, $\tilde{s}^0$ can be positive-definite in some frame with $\tilde{s}^\mu$ time–like.

In search of such a 4–current, Gerlach, Gromes & Petzold introduced a function, $F(p^\mu, p'^\mu)$ – called a 'form factor' –, into equation 9:

$$s^\mu = \frac{1}{(2\pi\hbar)^3}\int_{\mathbb{R}^6}\frac{d^3\mathbf{p}}{p_0}\frac{d^3\mathbf{p}'}{p'_0}\tilde{\psi}^*(\mathbf{p})\tilde{\psi}(\mathbf{p}')\exp\left(\frac{i(p_\mu - p'_\mu)x^\mu}{\hbar}\right)\frac{(p^\mu + p'^\mu)}{2}F(p^\mu, p'^\mu) \quad . \tag{6.16}$$

By considering the various constraints on $F$ imposed by the positivity of $\tilde{s}^0$ , covariance and normalization, they arrived at

$$F(p^\mu, p'^\mu) = \left(\frac{2m^2c^2}{p_\mu p'^\mu + m^2c^2}\right)^{1+\nu} \qquad (\nu \in \{1, 2, 3, \ldots\}) \quad . \tag{6.17}$$

I wish to make four points about this work:

First, it is not clear whether any further constraints could give a unique form factor, rather than the above, which is more accurately written with the functional dependence $F(p^\mu, p'^\mu, \nu)$ . It is to be expected of a formulation of the probability 4–current of a particle that it give a single expression. After all, it would be unusual for a particle to have two probability densities at a point, let alone a countable infinity of them. This is something Gerlach, Gromes & Petzold have not resolved.

The second problem is: the 4–currents, $\tilde{s}^\mu$, have all the right properties to be ideal for the job of describing the distribution and evolution of particle probability densities, except: the form factor has no acceptable physical interpretation, it is a mathematical artifice.

The third point against this development is, in my opinion, crucial. It is that the probability 4–current is not as fundamental a physical concept as the probability amplitude. There is no indication that any form of probability amplitude can be derived here. Without a probability amplitude over space-time (or something entirely equivalent to one) there is no basis for a quantum theory.

Finally, invariant probabilities are calculated from any 4–current by evaluating

$$\mathrm{pr}(H) = \int_H j^\nu \, d\sigma_\nu \tag{6.18}$$

in which $d\sigma^\nu = \left(dx^1dx^2dx^3, dx^0dx^2dx^3, dx^0dx^1dx^3, dx^0dx^1dx^2\right)$ , and $H$ is a space-like hypersurface – usually a constant-time hyperplane with respect to some coordinate system. The flaw in this is the dependence on the hypersurface, which makes the probability coordinate–dependent. Formula 18 may be the same in any inertial frame but $H$ is not. The probabilities are not acceptable on symmetry grounds, therefore, not because they are non-invariant but because for any fixed event, $(x)$ , there is no unique choice of surface, $H$ , passing through $(x)$ from which to calculate expectation values.

## §7 Position as a 4–vector of Operators

C. Dewdney, P. R. Holland, A. Kyprianidis, J. P. Vigier, *Relativistic Wigner Function As The Expectation Value Of The* PT *Operator*, Phys. Lett. **114A** (1986) 440–444.

J. R. Fanchi, W. J. Wilson, *Relativistic Many-Body Systems: Evolution Parameter Formalism*. Found. Phys. **13** (1983) 571–605.

L. P. Horwitz, F. Rohrlich, *Constraint Relativistic Quantum Dynamics*, Phys. Rev. **24D** (1981) 1528–1542.

L. P. Horwitz, C. Piron, *Relativistic Dynamics*, Helv. Phys. Acta **46** (1973) 316–326.

A. A. Broyles, *Space-Time Position Operators*, Phys. Rev. **1D** (1970) 979–988.

J. H. Cooke, *A Proper Time Formulation Of Quantum Mechanics*, Phys. Rev. **166** (1968) 1293–1298.

R. P. Feynman, *An Operator Calculus Having Applications In Quantum Electrodynamics*, Phys. Rev. **84** (1951) 108–128; notably Appendix D.

Y. Nambu, *The Use of the Proper Time in Quantum Electrodynamics (part I)*, Prog. Theor. Phys. **5** (1950) 82–94.

§§7.1 *Schrödinger wave mechanics with a Poincaré–Invariant Evolution Parameter*

The papers and authors listed above (excepting Broyles) are only a sampling of the efforts expended to produce a Poincaré–invariant analogue of Schrödinger's equation. By 'analogue' is meant a linear partial differential equation which is of first order in a single evolution parameter:

$$i\hbar\frac{\partial\psi}{\partial\tau} = \hat{H}\psi$$

On offer are, in the spin-0 case:-

$$i\hbar\frac{\partial\psi}{\partial\tau} = -\frac{\hbar^2}{2}\Box^2\psi$$

(Dewdney *et al.*, Cooke, Fock (according to Feynman), $\tau$ is conjugate to $m^2c^2/2$ ;

$$i\hbar\frac{\partial\psi}{\partial\tau} = -\frac{\hbar^2}{2mc}\Box^2\psi$$

(Horwitz, Piron, Fock (according to Nambu), Fanchi, Wilson), $\tau$ is conjugate to $mc/2$ ;

$$i\hbar\frac{\partial\psi}{\partial\tau} = \left(\hat{p}^\nu\hat{p}_\nu - m^2c^2\right)\psi$$

(Horwitz, Rohrlich), $\tau$ is conjugate to 0 ;

$$i\hbar\frac{\partial\psi}{\partial\tau} = \left(i\hbar\frac{\partial}{\partial t} - \hat{H}\right)\psi$$

(Nambu), $\tau$ is conjugate to 0 .

The last equation needs no further notice here as $\tau$ is not an invariant but is time-like.

Between the first and second equations there is only a multiplicative constant on the right. That said, the first is to be preferred and the second scorned: in both cases the rest mass becomes a quantum number, with $\tau$ as conjugate variable; if $m$ is no longer fixed it is not consistent to introduce a single rest mass as a multiplicative factor. Nevertheless, the dimensions of $\tau$ is not compatible with the interpretation of it as a 'proper time'. If $\tau$ were a proper time then it should have the dimensions of time, or perhaps length: as a variable conjugate to $m^2c^2/2$ , $\tau$ will be measured in units of $\mathrm{kg}^{-1}\mathrm{s}$ ; as the conjugate of $mc/2$ , the units are $\mathrm{m}^{-1}$ . This defect could be fixed if a new fundamental constant were used instead of $\hbar$ on the left-hand side of the dynamical equations.

The variability of the rest mass is itself a cause for concern. Two methods are employed to deal with this. The simplest ploy is to project onto a subspace of states with a single, definite value for the rest mass – this eliminates $\tau$ and recovers a more conventional wave equation: for spin-0, the Klein–Gordon equation. The second method, almost indistinguishable from no method at all, is to assume that the expectation value of the rest mass operator is the observed rest mass and that the distribution of values is so narrow that it is unnoticable.

Of course, where the variable $\tau$ remains a substantive part of the formalism, the resultant wave-function, $\psi = \psi(\tau, x)$ , will, generally, vary with $\tau$ ; i.e.,

$$\psi(\tau_1, x) \neq \psi(\tau_2, x)$$

– a feature bestowing on such theories the bizarre character of a variable history. Only James Cooke has an anywhere near believable explanation for such a corruption of fact. Cooke suggests that the wave-function describes an observation – extended over a region of space-time – and that the evolving wave-function then represents later observations. Provided such observations are bounded and strictly separated there would seem to be no problem. However, the evolution of the wave-function is continuous; so these 'observations', whether potential or actual, merge smoothly together. There is, further, no reason to suppose the wave-function to have bounded support on space-time for all 'times of observation', $\tau$ . For, finite support implies, via the Paley–Wiener theorem of Chapter 5, that the support in terms of the 4–momentum is analytic: not a condition particularly compatible with having very nearly an exact rest mass, nor with any assumptions such as Cooke makes:

$$p_0 > 0$$
$$p^\nu p_\nu > 0 \quad .$$

So either the interpretation of the theory in terms of observations is sensible, or the interpretation of the rest mass spectrum in the theory is straightforward. This suggests a new form of complementarity, or that the theory is untenable.

### §§7.2 *The Space-Time Representation: Hilbert Space and Normalisation*

Most of the authors listed at the head of this section are troubled by their invariant evolution parameters, particularly the way such parameters do not deserve to be called 'proper times' (see also the discussion of this in the work of Fleming (§6) and Derrick (Chapter 7)). The Hilbert space for the space-time representation provides another puzzle:

$$\mathcal{H} = L^2(\mathfrak{R}^4, d^4x)$$

If $\psi \in \mathcal{H}$ vanishes as $x^\nu \to \infty$ (for any $\nu$) then, surely, this means that the particle fades in and then out of existence? Admittedly this is over an infinite time-span, but it suggests strongly that the amount of particle is not a constant. Or, if it is required that the probability measure at any instant give unit probability then no physical state has a finite norm. Christopher Dewdney *et al.* adopted an idea by Olivier Costa de Beauregard by which normalisation is done only between two constant-time hyperplanes – allowing both

$$\int_{t_1}^{t_2} dx^0 \int |\psi|^2 \, d^3\mathbf{x}$$

and

$$\int |\psi|^2 \Big|_{x^0 \in [t_1, t_2]} d^3\mathbf{x}$$

to be finite. The advantage of this frame-dependent and cumbersome limiting hypothesis seems to me to be slight, at best.

The sole redeeming aspect of $\mathcal{H}$ is that the operators for position and momentum are simple, covariant 4–vectors of observables. The position observable is merely multiplication by $x$ . 4–momentum generates translations in 4–position, whence

$$[\hat{x}^\mu, \hat{p}_\nu] = i\hbar g^\mu_\nu \tag{7.1}$$

Of course, this means that position and rest mass are incompatible: if

$$\hat{m} = \hat{p}^\nu \hat{p}_\nu$$

then, it follows,

$$[\hat{x}^\mu, \hat{m}] = 2i\hbar g^{\mu\mu} \hat{p}_\mu \quad .$$

The Newton–Wigner operator does not suffer this 'defect', as a single rest mass is one assumption in its derivation; but then the components, insufficient in number as they are, can not be supplemented to satisfy equation 1. Equation 1 is, however, just the sort of condition a position observable might reasonably be expected to obey.

## §8 A Simple Conclusion

It is hardly surprising that theoretical interest has long since moved on, in the main, to quantum field theories and theories that avoid coordinate representations. A tenable, Poincaré–invariant quantum mechanics incorporating a credible rendition of locality has not been found in the sixty-two years since Schrödinger's first paper. It is just as well that gullible research students occur so frequently that such futile topics can be regularly shuffled about. The height of my ambition is to be counted among that happily frustrated throng.

## §9 Bibliography

§2 M. H. L. Pryce, *Commuting Co-ordinates in the New Field Theory*, Proc. Roy. Soc. **150A** (1935) 166–172.

M. H. L. Pryce, *The mass-centre in the restricted theory of relativity and its connexion with the quantum theory of elementary particles*, Proc. Roy. Soc. **195A** (1948) 62–81.

A. Papapetrou, *The concepts of angular momentum and the centre of gravity in relativity mechanics*, Praktika Akad. Athenon **14** (1939) 540–547. (See Appendix A for a translation of this paper.)

Chr. Møller, *On the definition of the centre of gravity of an arbitrary closed system in the theory of relativity*, Comm. Dublin Inst. Adv. Stud. ser. A no. 5 (1949).

Chr. Møller, *The Theory of Relativity*, O. U. P. 1972; pp 132–133, and pp 144–187.

§3 T. D. Newton, E. P. Wigner, *Localized states for elementary systems* Rev. Mod. Phys. **21** (1949) 400–406.

§5 G. N. Fleming, *Covariant position operators, spin, and locality*, Phys. Rev. **137B** (1965) 185–197.

§6 B. Gerlach, D. Gromes, J. Petzold, *The construction of definite expressions for the particle density of the Klein–Gordon field*, Z. Phys. **204** (1967) 1–11. (For a translation, see Appendix B.)

§7 C. Dewdney, P. R. Holland, A. Kyprianidis, J. P. Vigier, *Relativistic Wigner Function As The Expectation Value Of The* PT *Operator*, Phys. Lett. **114A** (1986) 440–444.

J. R. Fanchi, W. J. Wilson, *Relativistic Many-Body Systems: Evolution Parameter Formalism*, Found. Phys. **13** (1983) 571–605.

L. P. Horwitz, F. Rohrlich, *Constraint Relativistic Quantum Dynamics*, Phys. Rev. **24D** (1981) 1528–1542.

L. P. Horwitz, C. Piron, *Relativistic Dynamics*, Helv. Phys. Acta **46** (1973) 316–326.

A. A. Broyles, *Space-Time Position Operators*, Phys. Rev. **1D** (1970) 979–988.

J. H. Cooke, *A Proper Time Formulation Of Quantum Mechanics*, Phys. Rev. **166** (1968) 1293–1298.

R. P. Feynman, *An Operator Calculus Having Applications In Quantum Electrodynamics*, Phys. Rev. **84** (1951) 108–128; notably Appendix D.

Y. Nambu, *The Use of the Proper Time in Quantum Electrodynamics (part I)*, Prog. Theor. Phys. **5** (1950) 82–94.

(References are preceded with the page number on which they first appear.)

2: [1] A. J. Kálnay, *The Localisation Problem*, pp 93–110 in *Studies in the Foundations, Methodology and Philosophy of Physics*, vol. 4: *Problems in the Foundations of Physics*. Springer-Verlag 1971.

2: [2] A. S. Wightman, *On the localizability of quantum mechanical systems*. Rev. Mod. Phys. **34** (1962) 845–872.

2: [3] O. Costa de Beauregard, *On the zig-zagging causality EPR model: an answer to Vigier and co-workers and to Sutherland*, Found. Phys. **17** (1987) 775–785.

7: [4] E. P. Wigner, *On Unitary Representations of the Poincaré Group*, Ann. Math. **40** (1939) 149–204.

7: [5] S. Schweber, *An Introduction to Relativistic Quantum Field Theory*, 1961, p 55.

7: [6] V. Bargmann, E. P. Wigner, *Group Theoretical Discussion of Relativistic Wave Equations*, Proc. Nat. Acad. Sci. (U. S. A.) **34** (1956) 211–223.

10: [7] A. O. Barut, S. Malin, *Position operators and localizability of quantum systems described by finite- and infinite-dimensional wave equations*. Rev. Mod. Phys. **40** (1968) 632–651.

10: [8] E. P. Wigner, *Relativistic Invariance in Quantum Mechanics*. Nuovo Cimento **3** (1956) 517–532.

16: [9] D. I. Blokhintsev, *Space and Time in the Microworld*, D. Reidel 1973; p287.

16: [10] B. Gerlach, D. Gromes, J. Petzold, *A critical analysis of charge densities of Klein–Gordon fields*. Z. Phys. **202** (1967) 401–411.

16: [11] B. Gerlach, D. Gromes, J. Petzold, Z. Phys. **208** (1968) 381.

# Chapter 5

# The Anachoristic Theorem of Gerhard C. Hegerfeldt

*"I should have more faith', he said; 'I ought to know by this time that when a fact appears opposed to a long train of deduction it invariably proves to be capable of bearing some other interpretation.'*

**A Study In Scarlet** by Sir Arthur Conan Doyle.

## §1 Introduction

In 1974, Gerhard Hegerfeldt published a *No Go* theorem for what he termed a 'relativistic quantum theory', which I would dub a Poincaré–invariant quantum mechanics. With the aid of one of the many theorems on Fourier transforms elucidated by Raymond Paley and Norbert Wiener, Hegerfeldt produced a contradiction between an initially confined wave function and the Special Principle of Relativity. The impact of this is lessened when it is recalled that Poincaré–relativistic quantum mechanics already has a considerable number of fatal disfigurements: zitterbewegung, Klein's paradox, indefinite probability densities, too many position observables (or none at all),....[1] Another way of expressing Hegerfeldt's result is that limiting the theory to the positive–rest mass branch of the hyperboloid

$$p_\nu p^\nu = m^2 c^2 \tag{1.1}$$

destroys the hyperbolicity of the corresponding wave equation.

1976 found Bo–Sture Skagerstam attempting to prove Hegerfeldt's theorem using the Edge–of–the–wedge Theorem. Later, in 1977, Perez and Wilde published an article that utilised exactly the same line of reasoning. In fact, Skagerstam's result adds another facet to the picture begun with Hegerfeldt's theorem. It is shown that the class of wave functions chosen are non-zero almost everywhere in space-time.

By 1980 Hegerfeldt and Simon Ruijsenaars were trying (with, I suspect, little success) to improve on the generality of the first *No Go* theorem. More recently (1985) Hegerfeldt seems to have succeeded in relaxing the notion of localisation, though the linearity of the problem suggests that this was accomplished with more effort that was necessary. I will not discuss this last work here, as the previous results are quite enough to persuade me of the need for a more thorough analysis of Hegerfeldt's axioms. For there is nothing very new, original or controversial about these assumptions, yet they have quite unacceptable entailments.

The purpose of this chapter is to provide a straightforward review of this body of work. The complex analysis will be useful in the succeeding chapters, and indeed, it has already been alluded to in Chapter 3. A brief overview will be made of the theoretical problem revealed by this mathematics. The substantive pursuit of those possibilities that seem most promising will occupy the remainder of this work.

## §2 Some Complex Analysis

### §§2.1 *The Paley–Wiener Theorem*

An entire analytic function, $f(p)$ , is called an *exponential entire function* if, for some $A > 0$ ,

$$f(p) = O\left(e^{A|p|}\right)$$

i.e., $\exists k > 0$

$$\leq k e^{A|p|}$$

Now if $f(p) \in L^2(\Re)$ as a function of a real argument, then

$f(p)$ is an exponential entire function if an only if the Fourier transform $\tilde{f}(x)$ , of $f$ vanishes almost everywhere for $|x| > A$ .

Fortunately, it is only the sufficiency condition that will be used. The necessary condition turns out to be the difficult bit to show.

**Theorem** (Paley and Wiener, 1934[2])

If $\phi(x) \in L^2(\Re)$ vanishes almost everywhere for $|x| > A$ then

$$\tilde{\phi}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ipx} \phi(x)\, dx \tag{2.1}$$

is an exponential entire function.

**Proof**

Taking into account the support of $\phi$ , equation 1 is equivalent to

$$\tilde{\phi}(p) = \frac{1}{\sqrt{2\pi}} \int_{-A}^{A} e^{ipx} \phi(x)\, dx$$

whence

$$|\tilde{\phi}(p)| = \frac{1}{\sqrt{2\pi}} \left| \int_{-A}^{A} e^{ipx} \phi(x)\, dx \right|$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_{-A}^{A} e^{A|p|} |\phi(x)|\, dx$$

Since $|\exp(ipA)| \leq \exp(|A\,\Im(p)|) \leq \exp(A|p|)$ .

$$\leq \frac{\int_{-A}^{A} |\phi(x)|\, dx}{\sqrt{2\pi}} e^{A|p|}$$

Noting that $\tilde{\phi}(p)$ is entire completes the proof.

It should be remarked that the definition of exponential entire could be made more stringent by changing $\exp(A|p|)$ to $\exp(|A\,\Im(p)|)$ . The difference between these will not concern me except in showing that the canonical Schrödinger wave mechanics suffers 'instantaneous spreading' – a result of no shock value, and entirely compatible with invariance under the Galilei group.

This is not the most useful form of this mathematics. A more applicable form is obtained by taking what is called the *contrapositive* of the theorem. This is a matter of the symbolic logic of implications. Given the proposition

If $p$ then $q$ .

then the truth of this means that the proposition

If $q$ is not the case then $p$ is not the case.

is also true. The second sentence is the contrapositive of the first. A variety of proofs can be offered in the several forms of symbolic logic, of which perhaps the easiest is in terms of a boolean truth table:

| $p$ | $q$ | $p \Rightarrow q$ | not-$q \Rightarrow$ not-$p$ | not-$p$ | not-$q$ |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 0 | 0 |

The Paley–Wiener theorem above has the form of an implication (If ... then ...) whence the next theorem has also been proved.

**Theorem** (Contrapositive of the Paley–Wiener theorem)

If $\tilde{\phi}(p) \in L^2(\Re)$ is *not* an exponential entire function then its Fourier transform, $\phi(x)$ , has support with non-zero measure over an unbounded sub-set of $\Re$ .

§§2.2 *The Edge–of–the–wedge theorem*

The result used by Skagerstam and Perez & Wilde is called the Edge–of–the–wedge Theorem since, on the basis of a known function of a real variable, a function over complex values is deduced, e.g., for $\Re(z_1) > 0$ and $\Re(z_2) > 0$ –the real axis looks rather like the tip of a wedge in, in this instance, $\mathcal{C}^2$. The particular version of this theorem to be utilised is also known as the *Schwartz Reflection Principle*.[3]

Let $U$ be a domain in $\mathcal{C}$ that is symmetric about the real axis, so that for any $z \in U$ , then $\bar{z} \in U$ as well. Define three sub-sets:

$$U^+ = \{z \in U : \Im(z) > 0\}$$

$$U^- = \{z \in U : \Im(z) < 0\}$$

$$U^0 = \{z \in U : \Im(z) = 0\}$$

Take a function, $f : U^+ \cup U^0 \to C$ , to be continuous, analytic on $U^+$, and real-valued for $z \in U^0$; then there is an analytic function $F : U \to C$ that is an extension of $f$ to $U^-$.

$F$ is defined as

$$F(z) = \begin{cases} \overline{f(\bar{z})} & \text{, for } z \in U^- \ ; \\ f(z) & \text{, for } z \in U^0 \cup U^+ \ . \end{cases}$$

Now for any simple closed contour $\gamma$ in $U$

$$\int_\gamma F\,dz = \sum_j \int_{\gamma_j} f(z)\,dz + \sum_k \int_{\gamma_k} \overline{f(\bar{z})}\,dz$$

where the contours $\{\gamma_j\}$ all lie entirely in $U^0 \cup U^+$, and the contours $\{\gamma_k\}$ are all within $U^-$. Changing the variable in the second sum of integrals to $w = \bar{z}$ means that all the integrals are evaluated in $U^0 \cup U^+$. Since $f$ is analytic in $U^+$, and continuous on $U^0 \cup U^+$, Cauchy's theorem implies that all the integrals are zero. Thus

$$\int_\gamma F\,dz = 0 \quad .$$

This is true for any $\gamma$ in $U$ , and since $F$ is also continuous throughout $U$ , Morera's theorem[4] is applicable, viz. $F$ is analytic in $U$ .

Actually, it turns out that the Identity Theorem is also required for the conclusions drawn.[5]

**Theorem** (Identity)

If $f$ and $g$ are analytic in a domain $D$ and $f(z) = g(z)$ for all $z \in S \subseteq D$ where $S$ has a limit point in $D$ (a simple case being that $S$ is a line segment in $D$ ), then $f = g$ throughout $D$ .

This will be used with $f$ as the unknown function, $S$ some line or area of the complex plane and $g$ the zero function

$$g : z \to 0 \quad .$$

## §3 The *No Go* Theorems

### §§3.1 *Hegerfeldt's 1974 theorem*

G. C. Hegerfeldt, *Remark on Causality and Particle Localisation*, Phys. Rev. D **10** (1974) 3320—3321.

One of the most unpleasant things to find in an academic publication is the word 'obviously', simply because when an author feels it necessary to say that something is so transparent then there are going to be people to whom it is not. A more cunning ploy is to use a form of words which still means 'obviously' but without saying so. I mention this because, in his two-page article, Gerhard Hegerfeldt does just this, twice. The distressing aspect of the 'obviously' tactic is that it is very much more difficult to investigate when, as in this case, there is a crucial statement being proved. Given some background reading of Eugene Wigner's classic 1939 paper[6] and the additional clues provided in Skagerstam's proof, a respectable account of what is now called Hegerfeldt's Theorem might run as follows.

If $\phi$ is the state of a particle that definitely lies in some bounded spatial region, if $U(\mathbf{a})$ is the group of spatial translations, and if $U(t)$ is the group of time translations, then, in accordance with the Special Principle of Relativity, for any $t$ there ought to be an $r > 0$ such that

$$(\forall \mathbf{a} : |\mathbf{a}| > r) \qquad \langle U(\mathbf{a})\phi | U(t)\phi \rangle = 0 \quad .$$

Another way of putting this is: if a particle is once localised then it should remain confined within the light-cones having apices in that locality.

If it is now assumed that the particle is elementary and has an exact, positive rest-mass, then this equation can be re-written in terms of an irreducible unitary representation of the orthochronous

Poincaré group (*pace* Wigner). The limitation to the orthochronous sub-group is the implementation of the assumption that only positive 'relativistic energies' are admitted.

$$(\forall \mathbf{a} : |\mathbf{a}| > r(t))$$
$$\int_{\Re^3} \frac{d^3\mathbf{p}}{\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}} \sum_\sigma |\phi^\sigma(\mathbf{p})|^2 \exp\left(ict\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2} - i\mathbf{a}\cdot\mathbf{p}\right) = 0 \qquad (3.1)$$

The left-hand side is also the Fourier transform from **p**-space to **a**-space of

$$f(\mathbf{p},t) = \frac{\sum_\sigma |\phi^\sigma(\mathbf{p})|^2 \exp\left(ict\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}\right)}{\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}}$$

Equation 1 is therefore the mathematical statement that the Fourier transform, $\tilde{f}(\mathbf{a},t)$ , of $f(\mathbf{p},t)$ has its support entirely within the ball centred at $\mathbf{a}=\mathbf{0}$ with radius $r(t)$ . The second point to be noted is that, if $f(\mathbf{p},t)$ is analytic for some particular value of $t$ –e.g., zero, to simplify matters –, then the presence of the square root in the exponential means that this is not the case for all other values of $t$ . Or, rather, $f(\mathbf{p},t)$ is analytic for more than a single value of $t$ only if it is identically zero.

This implies that $f(\mathbf{p},t)$ can only be an exponential entire function (of $\mathbf{p}$) for one value of $t$ (zero) if it is non-trivial. For every other $t$ , by the contrapositive of the Paley–Wiener theorem, $\tilde{f}(\mathbf{a},t)$ does not have a bounded support. Equation 1 can, therefore, only be true for one, infinitesimal, instant. Alternatively, if an initial wave function of compact spatial support is assumed, then it will have propagated to an infinite support in the first instant of its evolution, i.e., moving at far in excess of the speed of light.

§§3.2 *Example: Schrödinger Wave Mechanics*

The received wisdom about the Schrödinger–Heisenberg quantum mechanics is that wave packets dissipate in time, spreading through space. The standard example is to calculate the evolution of a Gaussian. The contrapositive of the Paley–Wiener theorem now allows another simple demonstration of this aspect of the canonical quantum theory.

The arbitrary solution to the free Schrödinger equation can be written as

$$\psi(\mathbf{x},t) = \frac{1}{(2\pi\hbar)^{3/2}} \int_{\Re^3} d^3\mathbf{p} \exp\left(-\frac{i\mathbf{p}\cdot\mathbf{p}t}{2\hbar m} + \frac{i\mathbf{p}\cdot\mathbf{x}}{\hbar}\right) \tilde{\psi}(\mathbf{p})$$

where $\tilde{\psi}(\mathbf{p}) \in L^2(\Re^3, d^3\mathbf{p})$ . Now if $\psi(\mathbf{x},0)$ is a function of compact support it follows that $\tilde{\psi}(\mathbf{p})$ is an exponential entire function. Therefore, consider

$$f(\mathbf{p}) = \exp\left(-\frac{i\mathbf{p}\cdot\mathbf{p}t}{2\hbar m}\right) \tilde{\psi}(\mathbf{p})$$

for $\mathbf{p} = \mathbf{r} + i\mathbf{s} \in \mathcal{C}^3$. Now

$$\begin{aligned} |f(\mathbf{p})| &= \left|\exp\left(-\frac{it}{2\hbar m}(r^2 - s^2 + 2i\mathbf{r}\cdot\mathbf{s})\right)\right| \left|\tilde{\psi}(\mathbf{p})\right| \\ &= \exp\left(\frac{t}{\hbar m}\mathbf{r}\cdot\mathbf{s}\right) \left|\tilde{\psi}(\mathbf{p})\right| \end{aligned}$$

But this is not an exponential entire function (using the more stringent version of the exponential bound): it is analytic but is not bounded by

$$k \exp\left(\frac{|\mathbf{s}\cdot\mathbf{A}|}{\hbar}\right) \quad .$$

To see this, $\mathbf{r} \in \Re^3$ so, for example, take $\mathbf{r}$ parallel to $\mathbf{s}$ and of modulus greater than $Am/t$ .

Thus for any time other than the initial instant, $\psi(\mathbf{x},t)$ covers an unbounded subset of space.

§§3.3 *Skagerstam's Alternative Proof*

Bo–Sture K. Skagerstam, *Some Remarks Concerning the Question of Localisation of Elementary Particles*, Int. J. Theor. Phys. **15** (1976) 213—230.

J. F. Perez, I. F. Wilde, *Localisation and causality in relativistic quantum mechanics*, Phys. Rev. D **16** (1977) 315—7.

(This paper was submitted some months after the publication of Skagerstam's article.)

Repeating the assumptions made by Hegerfeldt in 1974, at greater length, Skagerstam arrives at equation 1 (his equation 3.3).

$$(3.1): \qquad \int_{\Re^3} \frac{d^3\mathbf{p}}{\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}} \sum_\sigma |\phi^\sigma(\mathbf{p})|^2 \exp\left(ict\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2} - i\mathbf{a}\cdot\mathbf{p}\right) = 0$$

Instead of examining the consequences of $\mathbf{p}$ being complex, Skagerstam chooses to consider the complexification of the 4–vector $(z) = (ct, \mathbf{a})$ . In fact, the left-hand side of equation 1 is well-defined on the wedge

$$W = \{\, z = x + iy : y^0 \geq 0, \quad y^\nu y_\nu \geq 0 \,\}$$

since the exponential becomes

$$\exp(ip_\nu z^\nu) = \exp(ip_\nu x^\nu)\exp(-p_\nu y^\nu)$$

and

$$\begin{aligned} -p_\nu y^\nu &= -y^0\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2} + \mathbf{y}\cdot\mathbf{p} \\ &\leq -y^0\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2} + |\mathbf{y}|\,|\mathbf{p}| \\ &\leq y^0\left(|\mathbf{p}| - \sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}\right) \leq 0 \end{aligned} \qquad (3.2)$$

So the presence of $y$ will only improve the convergence of the Fourier transform. Skagerstam's claim, however, is that

$$I(z) = \int_{\Re^3} \frac{d^3\mathbf{p}}{\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}} \sum_\sigma |\phi^\sigma(\mathbf{p})|^2 \exp\left(iz^0\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2} - i\mathbf{z}\cdot\mathbf{p}\right)$$

is not only properly defined on $W$ but that it is analytic with respect to $z$ off the real axis. Since

$$i\frac{\partial I}{\partial \mathbf{y}} = \frac{\partial I}{\partial \mathbf{x}} = \int_{\Re^3} -i\mathbf{p}\frac{d^3\mathbf{p}}{\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}} \sum_\sigma |\phi^\sigma(\mathbf{p})|^2 \exp\left(iz^0\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2} - i\mathbf{z}\cdot\mathbf{p}\right)$$

if the integral on the right exists then $I$ will be analytic; and since, for $|\mathbf{y}| > 0$ , there is a ball, $B(\mathbf{0}, R)$ so that for every $p \notin B(\mathbf{0}, R)$

$$|\mathbf{p}\exp(-p_\nu y^\nu)| < 1$$

it follows from the existence of the Fourier transform for $\mathbf{y} = \mathbf{0}$ (by the comparison test) that all derivatives of $I$ exist for $\mathbf{y} \neq \mathbf{0}$ . The only reason that the Schwartz reflection principle cannot be applied is now that the real function $I(x)$ is not known to be real-valued and continuous. This piece of the puzzle is provided by equation 1, at least for $|\mathbf{x}| \geq r$ .

Applying the version of the Schwartz Reflexion Principle on $C^4$ [7] to the left of equation 4 on the wedge

$$W_r = \{\, z = x + iy : |\mathbf{x}| \geq r(x^0), \quad y^0 > 0, \quad y^\nu y_\nu \geq 0 \,\}$$

gives a function analytic on this wedge and its reflection, that is zero on the subset of $\Re^4$ lying in the closure of this domain. By the Identity Theorem, $I(z) = 0$ on this domain. But $I$ is analytic throughout $W \backslash \Re^4$, so a further application of the Identity Theorem means that $I$ is zero on this subset of $W$ . Because $I$ is continuous on $W$ , it follows that $I(x)$ is zero for all 4–vectors $x \in \Re^4$.

This theorem provides essentially the same result as Hegerfeldt's 1974 theorem, but is not quite the same. Hegerfeldt showed that there were no states satisfying his assumptions that were spatially confined for more than one instant. Skagerstam has shown that there can be no space-time volumes (open subsets of $\Re^4$ ) on which, with the same assumptions, the wave function vanishes. Only Hegerfeldt's result implies superluminal propagation of the wave function; it is Skagerstam's result that the given assumptions mean that the wave function is non-zero almost everywhere in space-time.

### §§3.4 *Hegerfeldt and Ruijsenaars' 1980 'Generalisation'*

G. C. Hegerfeldt, S. N. M. Ruijsenaars, *Remarks on Causality, Localisation and Superreading of Wave Packets*, Phys. Rev. D **22** (1980) 377—384.

The 1980 collaboration between Gerhard Hegerfeldt and Simon Ruijsenaars is presented as an extension of previous work using somewhat less restrictive assumptions to obtain a further contradiction with the Special Principle of Relativity. The basis for this increment in generality is that the time evolution is now assumed merely to be generated by an operator with a semi-bounded spectrum (that is, a positive operator).

The first step is to prove a lemma, which is done using the Schwartz Reflexion Principle (cf. §2). Unlike the application of the edge–of–the–wedge theorem in the last sub-section, this result uses only a single complex variable. The conclusion is that if $U_t$ generates time evolution and $U_t\psi$ lies in a closed sub-space for $t \in [a,b]$ then (using the positivity of energy and the Reflection Principle) $U_t\psi$ lies in this subspace for all $t \in \Re$ . Of course, this does not exclude the trivial case in which the closed sub-space is the entire Hilbert space of states.

Two theorems are then laid out with which I would like to take issue. The problem with both is the use of a decomposition of the energy spectrum.

A state $\psi$ is assumed to be localised in some way. It is said to have an energy spectrum made up by the union of possibly an infinite number of finite, disjoint, intervals, $\{I_k\}$ . For each $I_k$ it is postulated that there is an open set of 3–momentum values, $O_k$ , which do not correspond to the energies in $I_k$ .

The localisation of $\psi$ is now expressed in the form familiar from the previous papers considered here: for $t \in [0,\epsilon]$ there is an $r$ such that

$$\langle U(\mathbf{a})\psi | U_t\psi\rangle = 0 \qquad (\forall \mathbf{a} : |\mathbf{a}| > r) \tag{3.3}$$

The authors' first theorem claims that $\psi = 0$ .

From the lemma, the range of times for which equation 3 is valid is actually $[-\infty,\infty]$ . The authors then make use of their decomposition of the energy spectrum:

> 'For any of the $I_k$ , let $\chi_{I_k}(p^0)$ be 1 on $I_k$ and zero outside. Then one also has
>
> $$\langle U(\mathbf{a})\psi | \chi_{I_k}(H)\psi\rangle = 0 \qquad (\forall \mathbf{a} : |\mathbf{a}| > r)' \tag{3.4}$$

This is apparently clarified by the footnote

> 'Note that $f(H) = \int \tilde{f}(t)\exp(iHt)\,dt$ , where the Fourier transform $\tilde{f}$ of $f$ is a continuous function.'

What this appears to mean is, as I see it, the following. Take a function $f(h) \in L^2(\Re)$ . The Fourier transform of this may be written

$$\tilde{f}(t) = \frac{1}{\sqrt{2\pi}} \int e^{-ith} f(h)\,dh$$

with inverse

$$f(h) = \frac{1}{\sqrt{2\pi}} \int e^{ith} \tilde{f}(t)\,dt$$

Now for any $f, g \in L^2(\Re)$

$$\widetilde{fg} = \tilde{f} * \tilde{g}$$

so that

$$\chi_I(h) f(h) = \frac{1}{\sqrt{2\pi}} \int e^{ith} \int \tilde{f}(t-s)\tilde{\chi}_I(s)\,ds\,dt$$

What Hegerfeldt and Ruijsenaars would like is for there to be some transform, $T$, so that

$$\chi_I(h) f(h) = \int T\left(\chi_{I_t}\right)(t)\, U_t f(h)\,dt \tag{3.5}$$

to give the locality property they desire. The problem is to, in some way, relate the right-hand sides of these two formulae. In the energy representation used here

$$U_t\psi = e^{ith}\psi(h)$$

$$\tilde{\chi}_{[a,b]}(t) = \frac{e^{-itb} - e^{-ita}}{-it}$$

So I can re-write the Fourier transform equation as

$$\chi_I(h)f(h) = \frac{1}{\sqrt{2\pi}} \int U_t \int \frac{1}{\sqrt{2\pi}} \int e^{-i(t-s)h'} f(h')\, dh'\, \tilde{\chi}_I(s)\, ds\, dt$$

This does not re-arrange into the form of equation 5.

I confess to being mystified by Hegerfeldt and Ruijsenaars' statement. As a first step the recondite definition of $\chi_I(H)$ can be recast in terms of spectral projectors (in the usual notation):

$$\chi_I(H) = E(\hat{H}; I) \quad .$$

Next, the assumption of a 'closed subspace' cannot be left as it is. This subspace must arise in some way; specifically, in some physically relevant way. Since $\psi$ is supposedly localised within a spatial volume, $V$ , for the time interval $[0, \epsilon]$ , the closed subspace must, therefore, have to do with this locality: where else does the orthogonality of $U(\mathbf{a})\psi$ derive? But if $\psi$ lies in a closed subspace that represents states localised on some particular volume, $V$ , of space, then there is a projector onto this subspace. I will denote this projector, for obvious reasons, by $E(\hat{x}; V)$ . There is no great difficulty in positing a complete set of projectors of this form – if not quite a resolution of the identity –, one for each volume, since I am not assuming that these operators are distinct or non-trivial. Now I can write the authors' assumption in the familiar form:

$$\psi = E(\hat{x}; V)\psi \quad .$$

The closed subspace is therefore the subspace

$$M = L^2(V) = E(\hat{x}; V)L^2(\Re^3) \quad .$$

The first conclusion drawn by Hegerfeldt and Ruijsenaars in the course of proving their theorem, by applying their first Lemma, can now be given the explicit, and potentially alarming form:

> If a state is confined to a volume $V$ for more than a single instant (or, 'a set of times of non-zero measure') then it will *always* be confined in that volume.

Thus I can already write

$$[E(\hat{x}; V), U_t] = 0 \quad ;$$

since the original assumption of a localisation in time has been conflated into an eternal localisation, so $U_t L^2(V) \subset L^2(V)$ ; and since, for any $\phi \in L^2(V)$ , I can write

$$\phi = U_t\, (U_{-t}\, \phi)$$

where I know that $U_{-t}\phi \in L^2(V)$ , so $L^2(V) \subset U_t L^2(V)$ .

It follows that 'strict localisation' in the form used here – the same form it takes in von Neumann's 1932 formulation of the Schrödinger–Heisenberg theory[5] – will only give a tenable theory if $E(\hat{x}; V) = \hat{I}$ , that is, $L^2(V) = L^2(\Re^3)$ ; or if $E(\hat{x}; V) = \hat{O}$ .

The gist of equation 4 is that $\chi_I(H)\psi$ has the same localisation as $\psi$ , i.e.,

$$E(\hat{H}; I)\psi = E(\hat{x}; V)E(\hat{H}; I)\psi \quad .$$

It is now possible to obtain this conclusion by cribbing a little from a theorem in Reed and Simon's textbook (M. Reed, B. Simon, *Methods of Modern Mathematical Physics*. A. P. 1972: p133, p272):

Let $f \in \mathcal{S}(\Re)$ ; then, by Fubini's theorem, for any $\phi$ and $\psi \in L^2(\Re^3)$ ,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(s)\langle e^{-is\hat{H}} E(\hat{x}; V)\phi, \psi\rangle\, ds &= \int_{-\infty}^{\infty} f(s) \left( \int_{-\infty}^{\infty} e^{-is\lambda}\, d_\lambda \langle E(\hat{H}; \lambda) E(\hat{x}; V)\phi, \psi\rangle \right) ds \\
&= \sqrt{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\lambda) d_\lambda \langle E(\hat{H}; \lambda) E(\hat{x}; V)\phi, \psi\rangle \\
&= \sqrt{2\pi} \langle E(\hat{x}; V)\phi, \tilde{f}(\hat{H})\psi\rangle
\end{aligned}$$

But, using the commutator established previously and then the same argument again,

$$\int_{-\infty}^{\infty} f(s)\langle e^{-is\hat{H}}E(\hat{x};V)\phi,\psi\rangle\,ds = \int_{-\infty}^{\infty} f(s)\langle e^{-is\hat{H}}\phi,E(\hat{x};V)\psi\rangle\,ds$$
$$= \sqrt{2\pi}\langle\phi,\tilde{f}(\hat{H})E(\hat{x};V)\psi\rangle$$
$$= \sqrt{2\pi}\langle E(\hat{x};V)\phi,\tilde{f}(\hat{H})\psi\rangle$$

In other 'words',

$$[E(\hat{x};V),f(\hat{H})] = 0 \tag{3.6}$$

But the characteristic function, $\chi_I(s)$ , can be expressed as the pointwise limit of a sequence, $\{f_n\}$ , of uniformly bounded functions in $\mathcal{S}(\Re)$ . Thus $f_n \to E(\hat{H};I)$ , and equation 6 holds throughout. Whence

$$[E(\hat{x};V),E(\hat{H};I_k)] = 0 \tag{3.7}$$

as required.

The remainder of the proof proceeds much as in Hegerfeldt's 1974 theorem, with the application of the contrapositive of the Paley–Wiener theorem. There is, to my mind, no very significant increment in the generality of this result over previous efforts. Only one spectral projector is assumed for position, but this was the case before. The use of a Hamiltonian with a semi-bound spectrum is the one novel feature, though what is gained it is hard to say – if there is no free-particle theory it seems improbable that there should be any other sort using the same assumptions. Skagerstam's proof can, in fact, be re-written in terms of a fairly general Hamiltonian: specifically, any function $H(\mathrm{p}) > |\mathrm{p}|$ .

## §4 The Paradox Considered

This is the paradox: by taking a small number of reasonable assumptions, a non-controversial consequence of Einstein's Special Principle is seen to be violated.

The assumptions made can be summarised as:

(i) that there is a class of states representing systems that are definitely confined to a bounded spatial volume;
(ii) that such states are orthogonal if they are space-like separated;
(iii) that space-like separations may be obtained by spatial translations (generated by the 3-momentum);
(iv) that these are states of positive energy from a unitary irreducible representation of the Poincaré group.

Abandoning or altering any one of the above brings with it problems that may be no smaller than the paradox to be averted. Tossing out axiom (i) is, I would hazard, the easiest; even though the formulation of a consistent theory is far from certain and there is always going to be the nagging worry of what all the absurdly small probabilities mean for systems that are as well confined as may be. There is also no guarantee that any alternative criterion, such as that devised by Gromes, Gerlach and Petzold (cf. the discussion of the work of these authors in chapter 4), will not be violated.

Discarding axiom (ii) would seem to be immediately daft. For non-orthogonality surely implies, as a transition amplitude, that the particle confined definitely in one volume might actually turn up in quite another. Further, drastic changes will be needed to make any sense of this choice.

To dispense with either assumption (iii) or (iv) is tantamount to renouncing the Special Principle – the equivalence of different reference frames –, a consequence hardly worth struggling towards. It may be possible to generate translations by some other means, though where this leaves the 3-momentum is uncertain. There is even the bizarre option that in translating a wave function the result may be a state that is no longer confined spatially, by means of a tortuous relation between the value of the wave function at each event and the probability of the presence of the particle. It may even be that some group other than the orthochronous Poincaré group is more condign.

**This is not a complete list, nor even a particularly well-balanced account, but it does suggest the scale of the problem. More detailed debate is postponed to the remaining chapters.**

## §5 Bibliography

**The variants on Hegerfeldt's theorem are to be found in the following papers, listed in chronoiogical order:**

G. C. Hegerfeldt, *Remark on Causality and Particle Localisation*, Phys. Rev. D **10** (1974) 3320—3321.

Bo–Sture K. Skagerstam, *Some Remarks Concerning the Question of Localisation of Elementary Particles*, Int. J. Theor. Phys. **15** (1976) 213—230.

J. F. Perez, I. F. Wilde, *Localisation and causality in relativistic quantum mechanics*, Phys. Rev. D **16** (1977) 315—7.

G. C. Hegerfeldt, S. N. M. Ruijsenaars, *Remarks on Causality, Localisation and Spreading of Wave Packets*, Phys. Rev. D **22** (1980) 377—384.

G. C. Hegerfeldt, *Violation of Causality in Relativistic Quantum Theory?* Phys. Rev. Lett. **54** (1985) 2395—2398.

---

(References are preceded by the page number on which they first occur.)

2: [1] S. S. Schweber *An Introduction to Relativistic Quantum Field Theory*, Row, Peterson & Co. 1961; *passim*.

2: [2] R. E. A. C. Paley, N. Wiener, *Fourier Transforms in the Complex Domain*, Amer. Math. Soc. Colloq. Pub. 19 (1934) p12.

This account is mainly derived from:

T. Kawata, *Fourier Analysis in Probability Theory*, A. P. 1972, pp 284–288.

3: [3] Ian Stewart, David Tall, *Complex Analysis*, C. U. P. 1983, p280.

4: [4] Stewart and Tall, *ibid*, p183.

4: [5] Stewart and Tall, *ibid*, p187.

4: [6] E. P. Wigner *On Unitary Irreducible Representations of the Inhomogeneous Lorentz Group*, Ann. Math. 40 (1939) 149–204.

6: [7] R. F. Streater, A. S. Wightman, *PCT, Spin & Statistics, and All That*, W. A. Benjamin Inc. 1964, Theorem 2–15 p77—82.

8: [8] J. von Neuman, *Mathematical Foundations of Quantum Mechanics*, O. U. P. 1955.

8: M. Reed, B. Simon, *Methods of Modern Mathematical Physics*, A. P. 1972; p133, p272.

# Chapter 6

## Localised Quantum Mechanics

*an exposition*

*'That's one small step for a man,*
*But a giant leap for a cripple.'*

**The Singing Detective** by Dennis Potter.

# §1 A Motivation

The ideas to be analysed here might be said to stem from the opening sentences of Rudolf Haag and Daniel Kastler's 1964 paper (*An Algebraic Approach To Quantum Field Theory*, J. Math. Phys. 3 848–61.)

> 'The essential feature which distinguishes quantum field theory within the frame of general quantum physics is the principle of locality. This principle states that it is meaningful to talk of observables which can be measured in a specific space-time region and that observables in causally disjoint regions are always compatible.'

One of the primary preoccupations of Wan and students (McKenna, McLean, Jackson, Timson)[1] has been to remove this unique attribute of quantum field theory by formulating a quantum mechanics that adheres to an identical Principle of Locality.

Of course, a new principle or axiom cannot merely be grafted onto quantum mechanics. For, not only is it possible that the new axiom may be contradictory, but it is also necessary to justify the addition – to offer a motivation for it. Should there be no good reason for an axiom then, by Occam's Razor, there is every reason to discard it. The advantage of a good motivation is, simply, that it builds an understanding of the physics into the mathematics at a fundamental level.

The motivation for the Principle of Locality is that quantum mechanics does not need to take the entire span of the universe into account every time an observation is made. There are certainly cosmologists and acolytes of 'wholism' who would not agree with this concept. However, since their theories – at least on this point – are of a metaphysical nature, rather than scientific, there is no onus to argue in favour of locality against such people.

Thus the extreme consequence of Schrödinger's wave mechanics that attracts almost all popular attention – that wave-functions are non-zero almost everywhere in the universe – is to be regarded as an 'edge effect' of the theory: at those places where the probability measure of a wave-function falls sufficiently close to zero it is not just unlikely that the particle will be found there but it will be taken to be impossible.

A more specific version of the Principle of Locality is:

(i) that all experimental determinations utilise apparatus of finite size, and so any representation of a measurement should employ a spatially confined object;

(ii) it is simply nonsensical for the particles involved in an experiment to be anywhere other than within the laboratory.

As an exegesis this is fine, but it is not sufficiently precise to use as a foundation for a formal treatment. The Principle of Locality enters the mathematical formalism in the definite form:

---

**Axiom** (*Principle of Locality*)
All observables and states representing real systems must have finite spatial support.

---

Any quantum theory based on this axiom will be called a *Localised Quantum Theory.*

There is an immediate consequence of this new premise, as Haag and Kastler pointed out, that seems fatal: the total energy, total charge, momentum, parity, and a number of other eminently physical quantities, cease to be observables. If the purpose of Wan *et al.* has been to introduce a principle of locality, the majority of their work has been to rescue and preserve the observable status of these quantities.

What follows (especially the next two sections) is a review of Dan Timson's attempt to produce a localised quantum mechanics which is invariant under the Galilei group.[1]

# §2 Quantum Mechanics Localised

The means by which a localised quantum mechanics is produced are the mappings called *localising isometries* and denoted $\mathrm{L}_\xi$. Each localising isometry provides a mapping of the 'global' states and observables of the canonical theory onto the localised states and observables which inform the localised theory. To proceed, the simplification of a single spatial dimension will be used.

A localising isometry is a mapping, $\mathrm{L}_\xi : L^2(\Re) \to L^2(\Re)$, on the Hilbert space $L^2(\Re)$, defined by

$$(\forall \psi \in L^2(\Re)) \qquad \mathrm{L}_\xi \psi = \tilde{\mathrm{L}}_\xi \oplus (0|_{\Lambda^c})\ \psi$$

where supp $\xi = \Lambda$ , $\Lambda^c = \Re\backslash\Lambda$ ; and $0|_{\Lambda^c}$ is the zero function on the set of points in $\Lambda^c$ .

Also

$$\tilde{L}_\xi : L^2(\Re) \to L^2(\Lambda)$$

is given by

$$\left(\tilde{L}_\xi\psi\right)(x) = \frac{\psi(\sigma(x))}{\sqrt{\xi(x)}} \qquad \text{if } x \in \Lambda \tag{2.1}$$

where

$$\sigma(x) = \int_{x_0}^{x} \frac{dy}{\xi(y)} + x_0 \quad , x_0 \in \Lambda_0 \quad , \tag{2.2}$$

and $\xi$ is assumed to be a positive, infinitely differentiable function (with, as mentioned before, support on $\Lambda$) that, for a closed sub-interval, $\Lambda_0 \subset \Lambda$ , is equal to one, which is the maximum value of $\xi$ . $\xi$ is called the *localising function.* $\Lambda_0$ is referred to as the *centre of localisation.*

$\sigma$ is a monotonic increasing function with range $\Re$ for the domain $\Lambda$ .

It follows from the foregoing that $\tilde{L}_\xi$ is a unitary and $L_\xi$ is an isometric transformation. The adjoint and inverse of $\tilde{L}_\xi$ is

$$\left(\tilde{L}_\xi^\dagger\psi\right)(x) = \psi(\sigma^{-1}(x))\sqrt{\xi\quad(\sigma^{-1}(x))} \quad . \tag{2.3}$$

This mathematics is now put to work in the first two axioms of the Timson–Wan theory:

---

**Axiom** (*States*)

At any instant of time the state of a quantum mechanical system will be an element, $\mathcal{H}$ , of a member, $S_\Lambda$ , of the family of augmented Hilbert spaces

$$S_\Lambda = \{\Omega : \Omega = (\psi_\Lambda, \xi_\Lambda), \quad \psi_\Lambda \in \mathcal{H}(\Lambda), \quad \Lambda \subseteq \Re\} \tag{2.4}$$

where each $\xi_\Lambda$ is a localising function such that the (Born) probability measure on the boundary of localisation is less than or equal to the significance level, $\epsilon$ :

$$\|E(\hat{x}; \Lambda_0)\psi_\Lambda\| \geq 1 - \epsilon \quad .$$

---

Given a state, and the localising isometry this implies, the next step is to produce an algebra of localised observables:

---

**Axiom** (*Observables*)

The *localised observable,* $\hat{A}_\Lambda$ , corresponding to a canonical observable, $\hat{A}$ , is defined by the formula

$$\hat{A}_\Lambda = L_\xi \hat{A} L_\xi^\dagger \quad . \tag{2.5}$$

---

$\hat{A}_\Lambda$ is a valid observable – it is self-adjoint – since $\hat{A}$ is a valid observable and $\tilde{L}_\xi$ is a unitary mapping. $\hat{A}_\Lambda$ also has the same spectrum as $\hat{A}$ , with the addition of zero if that value was not already in the spectrum of $\hat{A}$ . Since $\left(\tilde{L}_\xi\psi\right)(\Lambda_0) = \psi(\Lambda_0)$ , it follows that the localised observables are identical to the canonical, or 'global', observables on the, so-called, centre of localisation. The effect of this is that localised observables differ from their 'global' counterparts only in the part of the localisation that is not the centre of localisation ( $\Lambda\backslash\Lambda_0$ ); this region is called the *boundary of localisation.* It is the purpose of the axiom defining states to make this simple, local indistinguishability into a physically relevant property of the localised theory.

The complexity of these postulates arises from the more sophisticated connexion between a configuration space state, $(\psi_\Lambda, \xi_\Lambda)$ , and the corresponding localised momentum space state – not to mention all the other observables. The exact formulation is intended to ensure that where the 'global' observable and its localised counterpart differ there is not enough difference to worry about.

An example is the spectral projector for a general observable, $\hat{A}$ . If

$$E(\hat{A}; a)\psi = \psi$$

then

$$L_\xi E(\hat{A}; a) L_\xi^\dagger L_\xi \psi = L_\xi \psi$$

that is[2]

$$E(\hat{A}_\Lambda; a)\psi_\Lambda = \psi_\Lambda$$

so that

$$\left|\langle\psi_\Lambda \mid E(\hat{A};a) \mid \psi_\Lambda\rangle - \langle\psi_\Lambda \mid E(\hat{A}_\Lambda;a) \mid \psi_\Lambda\rangle\right| = \left|\langle\psi_\Lambda \mid E(\hat{A};a) - E(\hat{A}_\Lambda;a) \mid \psi_\Lambda\rangle\right|$$

(Applying the Cauchy–Schwartz inequality)

$$\leq \left\|E(\hat{A};a) - E(\hat{A}_\Lambda;a)\right\| \left\|E(\hat{x};\Re/\Lambda_0)\psi_\Lambda\right\|$$
$$\leq 2\epsilon$$

Taking $\epsilon$ to be sufficiently small means that, in this form, the 'global' and 'localised' versions of $\hat{A}$ are indistinguishable.

The easiest example for this localisation procedure is the standard example of an observable that does not commute with the position observable, viz. momentum.

Not all localised momentum ranges are necessarily compatible with every spatial localisation, given the form of the new axiom on states. It is at this point that Heisenberg's 'Uncertainty Principle' enters the theory. Or so it may seem.[3]

There are two formulations of this principle, which might be called 'uncertainty relations':

$$\Delta x \Delta p \geq \hbar/2 \tag{2.6}$$

and

$$[\hat{x},\hat{p}] = i\hbar\hat{I} \tag{2.7}$$

Of these, the first is the more commonly quoted, because it appears to have a straightforward meaning. That this is not so when one actually has a mathematical formalism to contend with must make this inequality a prime candidate for the title 'most popularly misunderstood formula in physics'.[4] Apart from anything else, this is an inequality of *variances*, which are statistical measures of only limited *physical* import, if any at all. If any characteristic volume can usefully be assigned to a wave packet it would be most reasonable to use the support of the function; though, of course, in Schrödinger's wave mechanics only one of $\psi(x)$ and $\tilde{\psi}(p)$ will have a finite support, if either has.

Since equation 6 is a consequence of equation 7, the latter might well be called the more fundamental version of Heisenberg's principle. The version to use with a localised evolution scheme may therefore be taken to be

$$[\hat{x},\hat{p}]\Big|_{\Lambda_0} = i\hbar\hat{I}\Big|_{\Lambda_0}$$

or

$$[\hat{x},\hat{p}_\Lambda] = i\hbar\xi\hat{I} \tag{2.8}$$

This is clearly a spatially dependent commutator in the boundary of localisation. The axiom defining states is sufficient to prevent any violation of equation 6. There is no conflict between this 'Uncertainty Principle' and a localised quantum mechanics.

## §3 Time Evolution Localised

To avoid the many contentious issues about Hamiltonians, only the free, Galilean, evolution will be discussed here:

$$H = \frac{p^2}{2m}$$

Naturally, this can be localised as a quantum mechanical observable by the procedure given in the last section.

$$\hat{H}_\Lambda = \mathrm{L}_\xi \hat{H} \mathrm{L}_\xi^\dagger = \mathrm{L}_\xi \frac{\hat{p}^2}{2m} \mathrm{L}_\xi^\dagger$$
$$= \frac{1}{2m} \mathrm{L}_\xi \hat{p} \mathrm{L}_\xi^\dagger \mathrm{L}_\xi \hat{p} \mathrm{L}_\xi^\dagger$$
$$= \frac{1}{2m} (\hat{p}_\Lambda)^2$$

– so the localisation process is consistent in this respect. There is, however, a definite problem with the time evolution operator, $\hat{U}(t)$ . The formula

$$\hat{U}_\Lambda(t) = \mathrm{L}_\xi \hat{U}(t) \mathrm{L}_\xi^\dagger$$

simply will not do: the support of a freely evolving state is not, generally, static. This lead Timson and Wan, in a attack of neologism, to devise what they called a *comoving evolution** – by which is meant that the localisation, $\Lambda$ (and, hence, the function $\xi_\Lambda$ ), as well as the wave-function, $\psi_\Lambda$ , evolves in time. This is formalised next.

---

**Axiom** (*Time Evolution*)
The time evolution of a state, $(\psi_\Lambda, \xi_\Lambda)$ , is generated by the Hamiltonian $\hat{H} = \hat{p}^2/2m$ according to the prescription

$$\psi_{\Lambda(t)}(t) = \mathrm{L}_{\xi(t)} \exp\left(\frac{i(t-s)\hat{p}^2}{2m\hbar}\right) \mathrm{L}^\dagger_{\xi(s)} \psi_{\Lambda(s)}(s)$$

where there is a bounded interval, $\Pi$ (the *momentum spectrum* of the state), such that,

$$\psi_{\Lambda(0)}(0) = E\left(\hat{p}_{\Lambda(0)}(0); \Pi\right) \psi_{\Lambda(0)}(0) \quad .$$

Also,

$$\underline{\Pi} = \inf\{\Pi\} \qquad \overline{\Pi} = \sup\{\Pi\}$$
$$\underline{\Lambda}(t) = \inf\{\Lambda(t)\} \qquad \overline{\Lambda}(t) = \sup\{\Lambda(t)\}$$

and

$$\overline{\Lambda}(t) = \overline{\Lambda}(s) + \frac{\overline{\Pi}}{m}(t-s)$$
$$\underline{\Lambda}(t) = \underline{\Lambda}(s) + \frac{\underline{\Pi}}{m}(t-s) \quad ;$$

and, repeating this prescription for the centre of localisation,

$$\overline{\Lambda}_0(t) = \overline{\Lambda}_0(s) + \frac{\overline{\Pi}}{m_0}(t-s)$$
$$\underline{\Lambda}_0(t) = \underline{\Lambda}_0(s) + \frac{\underline{\Pi}}{m_0}(t-s) \quad .$$

---

§§3.1 *Superposition of States*

So far I have re-written (in, I hope, a clear fashion) an account of the localised evolution of states of a fairly monolithic form: the support of the state in both configuration space and momentum space consists of a single, simply connected compact subset. Naturally, where a quantum theory has such 'monolithic states' one must also consider superpositions of such states. This goes beyond what Timson and Wan have developed.

If a monolithic state is produced by the successful passage of a particle through some preparing apparatus to its output, a superposition of such states arises from an apparatus with more than one distinct thoroughfare. One will then have an 'initial' (i.e., prepared) state, $\Phi$ , given by some prescription such as

$$\Phi = \alpha\phi_1 + \beta\phi_2 \tag{3.1}$$

( $|\alpha|^2 + |\beta|^2 = 1$ , $||\phi_i|| = 1$ ) where there are two phase space localisations ( $(\Lambda_1, \Pi_1)$ and $(\Lambda_2, \Pi_2)$ ) such that

$$E(\hat{x}; \Lambda_i)\phi_i = \phi_i$$
$$E(\hat{p}; \Pi_i)\phi_i = \phi_i \qquad (i = 1,2) \tag{3.2}$$

The two pairs of intervals need bear no especial relation with each other. For the 1–dimensional configuration space there are some thirty-six distinct cases – ignoring the values of the boundaries to the intervals but considering only the relative ordering of boundaries. It would be tedious to take each case in turn, so, for illustrative purposes, a suitably nasty example will be analysed in detail – see figure 1.

---

* Any student of English who, by whatever freakish chance, caught a glimpse of this term may be forgiven for feeling a bit faint. The absence of even a hyphen – to give 'co-moving' – is somewhat alarming.

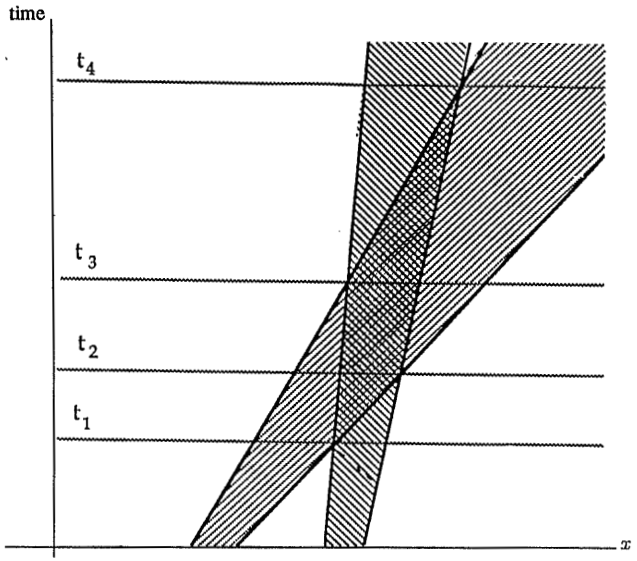



Figure 1: an unpleasant evolution.

§§3.2 *The Options*

The obvious candidates for the superposed time evolution are:

(i) $\phi_1$ and $\phi_2$ evolve by the localised evolutions they would each have if considered alone – adopting a fairly transparent notation for the modified time evolution operators, which are now functions of various localities, $\Lambda$ and $\Pi$ :

$$\Phi(t) = \alpha\hat{U}(t,\Lambda_1,\Pi_1)\phi_1 + \beta\hat{U}(t,\Lambda_2,\Pi_2)\phi_2 \quad ; \tag{3.3}$$

(ii) $\Phi$ evolves according to a law derived from the union of the two sets of ranges:

$$\Phi(t) = \hat{U}(t,\Lambda_1\cup\Lambda_2,\Pi_1\cup\Pi_2)\Phi(0) \quad ; \tag{3.4}$$

(see figure 2)

(iii) the composite evolution is based not on the unions of sets but on convex hulls, viz.

$$\begin{aligned}\Lambda &= \text{ConvexHull}(\Lambda_1,\Lambda_2)\\ \Pi &= \text{ConvexHull}(\Pi_1,\Pi_2)\end{aligned} \tag{3.5}$$

So that

$$\Phi(t) = \hat{U}(t,\Lambda,\Pi)\Phi(0) \tag{3.6}$$

(see figure 3).

By adding flesh to the rather skimpy options listed above, it will emerge that there is one that is naturally to be preferred.

Option (i) is already quite well defined, but there is a problem: for times between $t_1$ and $t_4$ the localised evolutions have over-lapping localisations. This puts paid to the unitary nature of the evolution since

$$\|\Phi(t)\|^2 = |\alpha|^2\|\phi_1\|^2 + |\beta|^2\|\phi_2\|^2 + 2\Re\big(\alpha\beta^*\langle\hat{U}(t,\Lambda_2,\Pi_2)\phi_2 \mid \hat{U}(t,\Lambda_1,\Pi_1)\phi_1\rangle\big) \tag{3.7}$$

$$= 1 + \text{ a non-zero term} \tag{3.8}$$

There is every reason to expect the cross-term to be significant.

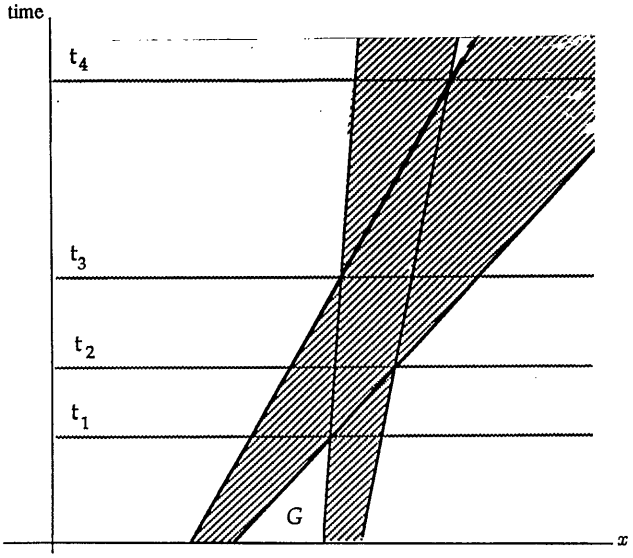



Figure 2: option (ii).

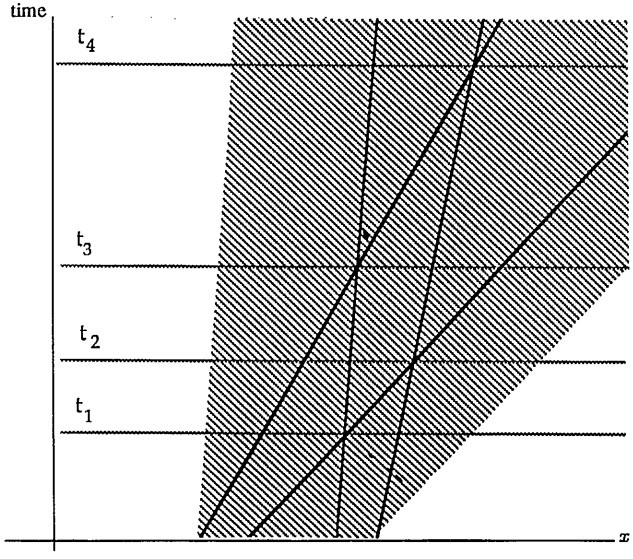



Figure 3: option (iii).

Option (ii): this is really just the natural result of noticing the flaws of the first guess. Here the evolution

would look, albeit at length, like the following.

$$\begin{aligned}
0 \le x^0 \le t_1 \ &: \Phi(x^0) = \alpha\hat{U}(x^0, \Lambda_1, \Pi_1)\phi_1 + \beta\hat{U}(x^0, \Lambda_2, \Pi_2)\phi_2 \\
t_1 \le x^0 \le t_2 \ &: \Phi(x^0) = \hat{U}\left(x^0 - t_1, \Lambda_1(x^0) \cup \Lambda_2(x^0), [\underline{\Pi}_1, \overline{\Pi}_2]\right)\Phi(t_1) \\
t_2 \le x^0 \le t_3 \ &: \Phi(x^0) = \hat{U}\left(x^0 - t_2, \Lambda_1(x^0) \cup \Lambda_2(x^0), [\underline{\Pi}_1, \overline{\Pi}_1]\right)\Phi(t_2) \\
t_3 \le x^0 \le t_4 \ &: \Phi(x^0) = \hat{U}\left(x^0 - t_3, \Lambda_1(x^0) \cup \Lambda_2(x^0), [\underline{\Pi}_2, \overline{\Pi}_1]\right)\Phi(t_3) \\
t_4 \le x^0 \ &: \Phi(x^0) = \left(\hat{U}(x^0 - t_4, \Lambda_1(x^0), \Pi_1) + \hat{U}(x^0 - t_4, \Lambda_2(x^0), \Pi_2)\right)\Phi(t_4)
\end{aligned} \tag{3.9}$$

This does not completely specify the evolution because the localising isometry is not determined in the region $\Lambda_1(x^0)\backslash\Lambda_{10}(x^0) \cap \Lambda_2(x^0)\backslash\Lambda_{20}(x^0)$ –the intersection of the two boundaries of localisation. There is no unique way of clearing up this lack of determinism, though there are plenty of rules that will work, for example

$$\xi(x) = \max\{\xi_1(x), \xi_2(x)\}$$

or

$$\xi(x) = \alpha\xi_1(x) + \beta\xi_2(x) \quad .$$

It is not difficult to see that this is unitary, and, in fact, respects very accurately the evolution of the component states. The major stumbling block for this option is that there is no obvious generalisation to two or more spatial dimensions. A less formal objection is that there is no particular reason why the region labelled $G$ should be left entirely untouched by the passage of the particle – it is the geometric shadow of $\Lambda_1$ and $\Lambda_2$ , but the particle has wave-like properties which do not necessarily exclude any such shadow.

Whilst the localised evolution can be said to hasten the process of asymptotic localisation (cf. Chapter 2), its main purpose is to provide a scheme in which position and momentum can be simultaneously confined for all times (finite as well as infinite). Thus, while the superposition consists of two states that are asymptotically separable, there is no reason to believe them separable much before $t_4$ .

This argument against option (ii) is further bolstered by the following observation on the extent of $\Lambda_1$ and $\Lambda_2$ : the limits of these localisations are established by much the same process in which statisticians set significance levels: a figure $(75\%, 90\%, 95\%, \ldots)$ is chosen as the bound on the accuracy of the model and/or experiment, and a centre of localisation ($\Lambda_{10}$ or $\Lambda_{20}$) is found which contains this proportion of the wave-function (using the Born measure). In this respect the boundary of at least the centre of localisation is somewhat arbitrary. Further, consider the theorem at the heart of asymptotic localisation:

$$\text{'If } E(\hat{\mathbf{p}}; [\underline{\Pi}, \overline{\Pi}])\Phi = \Phi \text{ then } \lim_{t\to\infty} E\left(\hat{\mathbf{x}}; \left[\frac{\underline{\Pi}}{m}t, \frac{\overline{\Pi}}{m}t\right]\right)\hat{U}(t)\Phi = \hat{U}(t)\Phi \quad .'$$

The space-time cone into which the particle finally falls has its vertex at the origin of coordinates, and this is regardless of how far any initial centre of localisation may be from the origin. It is true that the theorem can be readily adapted to a space-time cone with a vertex at any pre-arranged spot: a simple phase shift $\exp(i\hat{\mathbf{p}} \cdot \boldsymbol{\alpha})$ gives

$$\lim_{t\to\infty} E\left(\hat{\mathbf{x}}; \left[\frac{\underline{\Pi}}{m}t + \boldsymbol{\alpha}, \frac{\overline{\Pi}}{m}t + \boldsymbol{\alpha}\right]\right)\hat{U}(t)\Phi = \hat{U}(t)\Phi \tag{3.10}$$

and, indeed, convergence may be far more rapid for some values of $\boldsymbol{\alpha}$ than for others (since the evolution is a unitary automorphism, all the approximately localised states are already present in the Hilbert space, and there is no reason that an approximately localised state cannot remain so within some cone). What is clear is that the general statements of asymptotic localisation have little bearing on what occurs at small or just finite times.

§§3.3 *Option (iii): Superposition formulated*

Having, with somewhat greater reluctance, disposed of option (ii), I am left with option (iii). Here a consistently unified support has been found. There remains the problem of the local momentum operators. Now, for the separated component states there are localising isometries, call these $L_1$ and $L_2$ , such that

$$E(\hat{p}_{\Lambda_i}; \Pi_i)\phi_i = L_i E(\hat{p}; \Pi_i)L_i^\dagger\phi_i = \phi_i \qquad (i = 1, 2) \quad . \tag{3.11}$$

What is needed is a suitable isometry, L , so that

$$L\,E(\hat{p}; \Pi)L^\dagger\Phi = \Phi \quad . \tag{3.12}$$

Now $\Lambda$ is just another, quite unremarkable locality; so one can easily come up with a localising isometry for it, call this $L$ , for reasons that will become apparent. One might guess at $L$ having the form

$$L = L_1 + L_2 \quad . \tag{3.13}$$

However, between $x^0 = t_1$ and $x^0 = t_4$ there is, again, the problem of unitarity. In fact the spatial overlap rules out using $\phi_1$ and $\phi_2$ even as the initial components of $\Phi$ . It is here that the flexibility of the localisations $\Lambda_1$ and $\Lambda_2$ becomes actively beneficial; the $\{\phi_i\}$ can be re-localised onto $\Lambda$ with only a slight change of form:

$$\phi_i' = LL_i^\dagger \phi_i \qquad (i = 1,2) \quad . \tag{3.14}$$

This actually only alters the distribution of that part of the wave-function lying in the boundary of localisation, which was set up in the first place to be a less than significant fraction of the component state. This exercise also re-establishes the hegemony of the Schrödinger equation in the region $G$ . Thus the superposition may be freely taken to be

$$\Phi' = \alpha\phi_1' + \beta\phi_2' \tag{3.15}$$

which satisfies

$$\begin{aligned} &L\,E(\hat{p};\Pi)L^\dagger\Phi' = \Phi' \\ &L\,E(\hat{p};\Pi_i)L^\dagger\phi_i' = \phi_i' \qquad (i = 1,2) \quad . \end{aligned} \tag{3.16}$$

Establishing the support of the wave-function is one thing; producing a unique localising isometry is quite another. It is clear that a single localising isometry is required to supplant the isometries that apply only to one component, for it is only by having a single isometry that there is a set of observables, and thence an evolution of the system. Indeed, the principle requirement is to preserve, as much as possible, the confinement of the state to spatial and momentum volumes. By changing the estimate of the initial state using formula 14, the momentum spectrum of each component is retained exactly. To be more specific about the localising isometry, the only definite and obvious limitation is to take

$$\Lambda_0 = \text{ConvexHull}(\Lambda_{10}, \Lambda_{20}) \quad .$$

The choice of the remaining parts of the function $\xi$ is entirely open.

§§3.4 *Observables and the physical equivalence of states*

The discussion of re-localised states in the last section relies upon an assumption that will now be analysed. Obviously there is a difference between the state and its re-localisation, the question is not whether the two versions are mathematically indistinguishable, but whether they are physically so. The textbook definition of physical equivalence that might be thought applicable here is as follows:[5]

> 'Let $R_1$ and $R_2$ be two representations of the algebra $\mathcal{A}$ in the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ . These representations are called *physically equivalent* if for any choice of the positive integer $n$ and observables $A_1, \ldots, A_n \in \mathcal{A}$ , any positive trace-class operator $\hat{B}_1$ in $\mathcal{H}_1$ , and any arbitrarily small (fixed) $\epsilon > 0$ , there exists a positive, trace-class operator $\hat{B}_2$ in $\mathcal{H}_2$ such that
>
> $$|\text{Tr}(\hat{B}_1 R_1(A_k)) - \text{Tr}(\hat{B}_2 R_2(A_k))| < \epsilon \qquad k = 1, \ldots, n \quad .' \tag{3.17}$$

This is not, in fact, a very 'physical' definition at all. The following, slight, adaptation is in accordance with the concept of dimensional analysis in physics, i.e., it accounts for the fact that $\langle\psi|\hat{A}|\psi\rangle$ is an expectation value of an observable, such as momentum, whereas the error level, $\epsilon$ , will be dimensionless. So, instead of equation 17, I propose to use

$$\frac{|\text{Tr}(\hat{B}_1 R_1(A_k)) - \text{Tr}(\hat{B}_2 R_2(A_k))|}{|\text{Tr}(\hat{B}_1 R_1(A_k))|} < \epsilon \qquad k = 1, \ldots, n \quad . \tag{3.17'}$$

Clearly systems equivalent according to equation 17 are equivalent according to equation 17′ , however the reverse is not true. The importance of the change is in the application of the revised definition in the present case. Consider the innocuous-seeming observable

$$\bar{A} = \gamma E(\hat{\mathbf{x}}, \Lambda_g) \tag{3.18}$$

where $\Lambda_g \subset G$, and $|\gamma| \in \Re$. The problem is that

$$\begin{aligned} \bar{A}\phi_i &= 0 \\ \bar{A}\phi_i' &\neq 0 \end{aligned} \tag{3.19}$$

so that, using equation 17, and if $|\gamma|$ is great enough, one could easily get

$$\left| \|\bar{A}\phi_i\| - \|\bar{A}\phi_i'\| \right| > \epsilon \tag{3.20}$$

for any $\epsilon$. This problem does not arise for equation 17′, where the left-hand side is divided by a factor proportional to $|\gamma|$. Since it is precisely in the region $G$ that differences are manifest, using equation 17′ ensures that these differences between $\phi_i$ and $\phi_i'$ are of the same order as

$$\|E(\hat{\mathbf{x}}; \Lambda_i^c)\phi_i'\| \tag{3.21}$$

which is, by assumption, a very small quantity.

# §4 Symmetry: 'Global' v. 'Local'

In performing experiments at different times and places, experimenters have always assumed that the natural laws being studied do not change despite the differing circumstances. From such humble beginnings, the concept of symmetry has become central within all physical theories. The feature of the realisation of the symmetry group in the canonical quantum mechanics that is not in keeping with a localised theory is the global action of the group representation. A localised notion of symmetry would seem to be in order.

Localised or not, the representation of symmetry in a quantum theory is still a matter of satisfying the criterion eloquently enunciated by Wigner in 1939. If $\hat{U}$ is a symmetry operator then all probabilities are preserved:

$$(\forall \psi, \phi \in \mathcal{H}) \qquad |\langle \psi \mid \phi \rangle| = \left| \langle \hat{U}\psi \mid \hat{U}\phi \rangle \right| \tag{4.1}$$

This implies that $\hat{U}$ is either a unitary or an anti-unitary transformation. If $\hat{U}$ is to be an element of a simple continuous group it follows that it is unitary; a state of affairs that will be assumed to apply in the sequel.

It is not unusual to suggest that this is the fundamental symmetry requirement but that further demands may be made of a theory under the heading of *'manifest invariance'* (I do not claim to have found the original or best expositions of this notion, but the two papers by Foldy, on the one hand, and Currie, Jordan and Sudarshan, on the other, are as good as I have seen on this point.[6][7]) The operator equivalent of the (Poincaré) covariant transformation of coordinates falls in this second category, and not least because of the difficulty authors have had in devising a credible position observable.

That the concept of 'manifest invariance' is a fudge is a conclusion I find reasonable on the following grounds: if there is a position observable, $\hat{x}$, then equation 1 may be re-written, with $\psi \in \mathcal{D}(\hat{x})$ and $\phi = \hat{x}\psi$, as

$$\begin{aligned} |\langle \psi \mid \hat{x}\psi \rangle| &= \left| \langle \hat{U}\psi \mid \hat{U}\hat{x}\psi \rangle \right| \\ &= \left| \langle \hat{U}\psi \mid (\hat{U}\hat{x}\hat{U}^\dagger)\,\hat{U}\psi \rangle \right| \end{aligned}$$

in other words, given the 'fundamental' requirement of symmetry, I deduce that the operator transformation law for position (and, by extension, any observable) is

$$\hat{x}' = \hat{U}\hat{x}\hat{U}^\dagger \tag{4.2}$$

with a corresponding transformation of expectation values. Now if $\hat{U}$ represents some straightforward symmetry – translation, rotation, Galilean boost, Lorentz boost – then it is no more than sensible to insist that the operator transformation law be a transcription of the classical covariance statement.

On the contrary, if there were, for whatever reason, no position observable actually worthy of the name then one way of avoiding the painful admission that the whole edifice is junk is to invent the spurious distinction that 'manifest invariance' is optional. After all, a theory that is 'not manifestly covariant' sounds much more promising than a theory that incorporates no sensible notion of space.

§§4.1 *Localised Symmetry, marque 1*

If the canonical generator of a canonical, 'global', symmetry is the observable $\hat{g}$ then the obvious localisation of symmetry is to use

$$\hat{g}_\Lambda = L_\xi \hat{g} L_\xi^\dagger \tag{4.3}$$

to generate, by way of the localised version of Stone's theorem, the group of operators

$$\hat{U}_\Lambda(s) = L_\xi \exp\left(\frac{\hat{g}s}{i\hbar}\right) L_\xi^\dagger \tag{4.4}$$

There are two reasons why this naïve formula is untenable: (i) in general, the axiom defining states is violated; and (ii) if the untransformed state lies in some representation then the transformed state does not. An example will illustrate the problem here, for there is only one, at base.

Take a state, $\psi$, localised to the interval $[a, b]$, with centre of localisation $(a_0, b_0)$. Consider $\psi$ translated by

$$s = \frac{b_0 - a_0}{1000}, \quad \frac{b_0 - a_0}{100}, \quad \frac{b_0 - a_0}{10}, \quad b_0 - a_0, \quad 10(b_0 - a_0)$$

Tracing the action of the operations in formula 4 in order, first of all a state in $L^2(\Re)$ is produced (identical to $\psi$ on $(a_0, b_0)$). The conventional translation then shunts this curve along the spatial axis, so the values on $(a_0, b_0)$ are now taken on $(a_0 + s, b_0 + s)$. Finally, the localising isometry $L_\xi$ brings the state back to its original support. Taking the list of shifts in order, the result of equation 4 can be assessed. For the first and second values of $s$ there is probably no difficulty; there may even be no cause for concern over the third value. The later values in the list will, it can not be doubted, be highly disruptive. For $s = b_0 - a_0$ the centre of localisation is pushed entirely into a boundary region, with the immediate consequence that the Born measure in this boundary is now quite close to unity, in direct conflict with the axiom defining states.

Another, perhaps more alarming, way of putting this is to consider the expectation value of position.

Before: $$\langle x \rangle = \langle \psi \mid \hat{x}_\Lambda \mid \psi \rangle = \langle L_\xi^\dagger \psi \mid \hat{x} \mid L_\xi^\dagger \psi \rangle$$

After: $$\begin{aligned} \langle x' \rangle &= \langle \hat{U}_\Lambda \psi \mid \hat{x}_\Lambda \mid \hat{U}_\Lambda \psi \rangle \\ &= \langle L_\xi^\dagger \psi \mid \hat{x} - s \mid L_\xi^\dagger \psi \rangle \\ &= \langle x \rangle - s \quad . \end{aligned}$$

So, if $\langle x \rangle = a_0 + \alpha(b_0 - a_0)$ then $\langle x' \rangle = a_0 + \alpha(b_0 - a_0) + s$, and hence, for $s > b_0 - a_0$, $\langle x' \rangle$ will be outwith the localisation! It was for this sort of situation that the axiom defining states takes the form it does.

§§4.2 *Localised Symmetry, marque 2*

A workable scheme can be had for only a modest revision of formula 4. Instead of merely transforming $L_\xi^\dagger \psi$, the function $\xi$ is also transformed, giving $\zeta$, say, in the final localising isometry. Whence $\hat{U}_\Lambda(\xi_\Lambda, s)$ : $(\psi_\Lambda, \xi_\Lambda) \mapsto (\psi'_{\Lambda'}, \zeta_{\Lambda'})$ by

$$\hat{U}_\Lambda(\xi_\Lambda, s) = L_\zeta \exp\left(\frac{\hat{g}s}{i\hbar}\right) L_\xi^\dagger \tag{4.5}$$

More precisely, if there is a representation of the continuous group that acts on the points of space by

$$x' = G(s)x$$

then

$$\left(\hat{U}_\Lambda(\xi_\Lambda, s)\psi_\Lambda\right)(x) = L\left(\xi(G^{-1}(s)x)\right) \exp\left(\frac{\hat{g}s}{i\hbar}\right) L^\dagger(\xi(x))\psi_\Lambda(x) \tag{4.6}$$

There is a group property if $\xi'(x) = \xi\left(G^{-1}(s)(x)\right)$ (i.e., $\xi'$ is the transform of $\xi$) in the process of successive transformations: $\hat{U}_{\Lambda'}(\xi'_{\Lambda'}, t)\hat{U}_\Lambda(\xi_\Lambda, s)$

i.e.,

$$\begin{aligned} \left(\hat{U}_{\Lambda'}(\xi'_{\Lambda'}, t)\hat{U}_\Lambda(\xi_\Lambda, s)\psi_\Lambda(x)\right)(x) = L\left(\xi(G^{-1}(t)G^{-1}(s)x)\right) \exp\left(\frac{\hat{g}t}{i\hbar}\right) L^\dagger\left(\xi(G^{-1}(s)x)\right) \\ L\left(\xi(G^{-1}(s)x)\right) \exp\left(\frac{\hat{g}s}{i\hbar}\right) L^\dagger(\xi(x))\psi_\Lambda(x) \end{aligned} \tag{4.7}$$

In point of fact, $\hat{U}_\Lambda(\xi_\Lambda, s)$ is indistinguishable from $\hat{U}(s) = \exp(\hat{g}s/i\hbar)$ . After all,

$$\exp\left(\frac{\hat{g}s}{i\hbar}\right) \mathrm{L}^\dagger(\xi(x))\psi_\Lambda(x) = \mathrm{L}^\dagger(\xi(G^{-1}(s)x))\psi_\Lambda\left(G^{-1}(s)x\right)$$

whence

$$\begin{aligned}\left(\hat{U}_\Lambda(\xi_\Lambda, s)\psi_\Lambda\right)(x) &= \psi_\Lambda\left(G^{-1}(s)x\right) \\ &= \left(\exp\left(\frac{\hat{g}s}{i\hbar}\right)\psi_\Lambda\right)(x) \quad .\end{aligned}$$

The transformation of $\xi$ also means that the symmetry group acts in the same way as time evolution: changing the whole state, $(\psi_\Lambda, \xi_\Lambda)$, and not just the wave-function, $\psi_\Lambda$ .

# §5 An analysis of concepts

§§5.1 *Observables and Symmetries*

A reasonable facsimile of the usual quantum mechanics has been sketched in the preceding sections, but with the added bonus that there is no need to look beyond a finite spatial volume. The task of implementing the Principle of Locality might well seem near achievement. On the other hand, the final result of the last section was that the localisation region, $\Lambda$ , was not the largest volume necessary for a quantum mechanics localised to that localisation region. If the generator of a symmetry is a localised observable, it must be concluded that its 'centre of localisation' contains $\Lambda$ (which might still be called the localisation *of the particle,* I suppose) as a proper subset.

Timson and Wan have argued that *all* observables on a localised quantum mechanics should have $\Lambda$ as their support.[2] Their reasoning was that for a smaller support the observable would not be 'effective' – would not be certain to measure the particle – whereas having a larger support would be indistinguishable from having $\Lambda$ for a support. This simple-minded explanation is, if not wrong then, surely, it is dubious. The explanation is necessitated by the fact that this is a theory in which only one locality can be used at a time: the localisation of the state and the support of the localised observables is always identical – localised observables not defined with respect to the localising isometry of the state simply mean nothing.

There is no fault in supposing that a measuring device can have a smaller support than the wave-function of the object it measures. There has been no bound put on the size of $\Lambda$ , on the one hand, and on the other, the miniaturisation of detectors is quite advanced by now.

To argue that the smaller detectors may miss the object is to forget that even devices very much larger than any sensible particle support have efficiencies of rather less than one – in the terminology of Clauser *et al.*[8] A good example is the gigantic size of neutrino detectors.

It is as untenable to hold that localised observables have a support no larger than $\Lambda$. Again, neutrino detectors and cloud chambers do not work on this principle: the design of detectors has never been a matter of matching exactly the supposed support of a wave-function. The fact that the support of an observable that 'generates' a symmetry is always larger than $\Lambda$ is quite telling, if there is still good reason to localise it. To obtain a localised 'generator' it is necessary to have as its centre of localisation, at the minimum, the localities $\Lambda$ and the transformation of $\Lambda$ ($G(s)\Lambda$ in the symbols of the last section). Thence it is only possible to localise a 'generator' for a bounded range of the transformations it is to generate.

If the symmetry group is intended to exhibit the degree to which the physics is 'coordinate free' (for the idealised space-time adopted) then there is no reason to limit the range of transformations; so the 'global' symmetry group is the right one after all.

It seems that the support of the localised observables and the spatial extent of the measuring devices that these are supposed to represent are unconnected.

§§5.2 *The nature of* $\xi$

So what, I am lead to inquire, is the nature of the function $\xi$ ? What does it represent? If it does not represent the 'footprint' of measuring devices, then it must only be associated with the quantum object. If

the support and 'centre' of $\xi$ give the localisation of a particle and the localisation of the 'bulk' of a particle, respectively, what then of the infinite variability of $\xi$ ? Given a $\Lambda$ and a $\Lambda_0 \subset \Lambda$ , there is no 'natural' or 'obvious' candidate function, $\xi$ . Admittedly, the various allowable functions only differ on the 'boundary of localisation', $\Lambda \setminus \Lambda_0$ , where such differences are claimed to have no noticeable effect.

To make this quite clear, take a wave-function of compact support, $\Lambda$ . Where is the centre of localisation of this wave-function? As was mentioned in connection with the superposition of states (§§3.2), having set the significance level there is no natural, that is, unique, boundary within $\Lambda$ to exclude only the 'insignificant' part of the state.

Given a set of bounded observables, $\{ \hat{B}_i \}$ , where

$$b = \max_i \{ \|\hat{B}_i\| \}$$

then by setting the significance level, $\epsilon$ , in the axiom on states to be $\delta/2b$ , where $\delta$ is the largest measurable difference (the accuracy of the detectors), it follows from the result of §2 that there will be no way of distinguishing between the competing forms of $\xi$ .

If $\psi_1 = L_{\xi_1}\psi$ and $\psi_2 = L_{\xi_2}\psi$ , wherein the supports and centres of $\xi_1$ and $\xi_2$ are identical, then

$$\begin{aligned} &\left|\langle\psi_1 \mid \hat{B} \mid \psi_1\rangle - \langle\psi_2 \mid \hat{B} \mid \psi_2\rangle\right| \\ &\qquad = \left|\langle\psi_1 \mid E(\hat{x};\Lambda\setminus\Lambda_0)\hat{B}E(\hat{x};\Lambda\setminus\Lambda_0) \mid \psi_1\rangle - \langle\psi_2 \mid E(\hat{x};\Lambda\setminus\Lambda_0)\hat{B}E(\hat{x};\Lambda\setminus\Lambda_0) \mid \psi_2\rangle\right| \\ &\qquad \leq \left(\|E(\hat{x};\Lambda\setminus\Lambda_0)\psi_1\| + \|E(\hat{x};\Lambda\setminus\Lambda_0)\psi_2\|\right)\|\hat{B}\| \leq 2\epsilon b \end{aligned}$$

This is, to my mind, not the right way around. To produce a testable theory it is not necessary to take into account the fallibility and finite engineering of any experimental use of the theory. The fact that experiments are always subject to error is a matter for the experimenter and philosophers, not theorists. The approach reviewed here, like Eduard Prugovečki's 'Stochastic Quantum Mechanics' (D. Reidel, 1985), uses as one of its premises the accuracy of detecting equipment. The theory therefore adapts itself to differing experiments; so much so I begin to doubt that it is refutable. It is not a question of which $\xi$ forms part of an objective, physical, description – and so which $\xi$'s do not – but of finding a $\xi$ to fit the facts.

§§5.3 *Initial Conditions : Initial States*

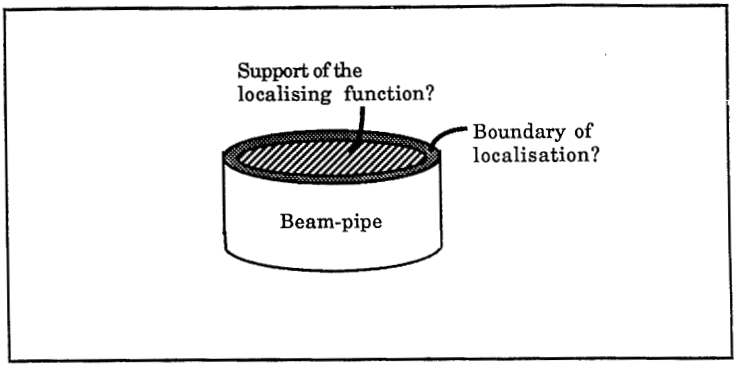
To specify a problem in the localised quantum theory discussed here requires the same thing as in the standard Galilei–invariant theory: the initial state must be found. The difference is that in the localised theory the states – initial or otherwise – consist of a wave-function and a localising function: $(\psi_\Lambda, \xi_\Lambda)$ . Further, the localising function, through the definition of a local momentum operator, is also crucial in determining the time evolution. Therefore the specification of a particular evolution in the localised quantum theory poses greater demands.

Consider, as a rough but workable guide, the process by which an initial state might be established in each theory. It may be supposed the quantum system enters the experiment through a beam-pipe. Now if the initial state is to allow the subsequent behaviour of the system to be accurately predicted, it would be prudent to measure as precisely as possible the state of the systems emerging from this pipe. The unobservability of the wave-function is a burden shared by both localised and canonical quantum mechanics; necessitating an indirect form of deduction.

A simple measurement that gets quite close to the wave-function is to place a photographic plate on the end of the beam-pipe – giving the probability density, $|\psi(x)|^2$ . This is liable to be close enough for a good estimate of the initial state in the canonical theory. The position measurement, despite the compatibility of the empirical error and theoretical approximation, is not sufficient for the localised theory. By the use of a uniform magnetic field (for charged particles, at least) and another photographic plate, the probability density, $|\tilde{\psi}(p)|^2$ , can also be measured. With this second piece of information it may now be possible to estimate an initial state for the localised theory. The support of $\xi$ and the centre of localisation will presumably be found by measurement of the end of the pipe; the actual form of $\psi$ and $\xi$ will be found by reconciling the observed momentum spectrum and $|\psi(x)|^2$ –though now the edges of this distribution will be treated as less than trustworthy.

For the canonical theory, of course, the measurement of these two probability densities represents a test of the theory – if not necessarily an exacting one. For the localised theory, there is no guarantee that the attempted reconciliation of the measurements – despite the infinite variability of $\xi$ –will give an initial state; and so, to this extent, the theory is also tested.

Removing these measuring devices, the quantum system, as it now enters the experimental region proper, can be said to be 'prepared' with the distributions of dynamical variables that have been found. At this initial




Estimating the localising function, $\xi$ .

instant the localisation of particle and observables can be convincingly said to coincide. For subsequent times the momentum observable must be $\hat{p}_\Lambda$ , though the support of the observable and the active part of a momentum measuring device will be unrelated. What is true of momentum will also be true of other measurable quantities.

Finally, if there is more than one beam-pipe feeding into the experiment then the superposed state must be formed by the method described in §3, above. This adaptation of the estimate of the initial state will not be noticed among the other approximations in use.

§§5.4 *Time Evolution*

If the localisation of states and observables is *ad hoc*, then the, so-called, comoving time evolution goes further. In redefining the evolution in time a stronger than usual interpretation is laid on the momentum spectrum, viz. that this allows a spray of particle trajectories to be plotted. There is, however, a problem in showing the consistency of the axioms defining states and time evolutions.

Given $\psi \in \mathcal{H}(\Lambda)$ , with local momentum operator $\hat{p}_\Lambda$ , such that

$$\|E(\hat{x};\Lambda_0^c)\psi\| < \epsilon'$$
$$E(\hat{p}_\Lambda;\Pi)\psi = \psi$$

then

$(\forall \epsilon > \epsilon')(\exists t_1, t_2)$ such that $(\forall t \in [0,t_1) \cup (t_2,\infty))$

$$\left\|\left(\hat{U}_{\Lambda(t)}(t) - \hat{U}(t)\right)\psi\right\| < \epsilon \quad .$$

The time $t_1$ exists from the definition of the localised evolution:

$$\hat{U}_{\Lambda(t)}(t) = \mathrm{L}_{\xi(t)}\hat{U}(t)\mathrm{L}^\dagger_{\xi(0)}$$

and the strong continuity of $\hat{U}(t)$ ( $\mathrm{L}_{\xi(t)}$ is assumed to be strongly continuous in $t$ ). $t_2$ exists by the theorem on asymptotic localisation.

Setting aside the sizable task of finding these times, they can be assumed to be ordered $t_1 < t_2$ . Between these instants there is no mathematical reason (that I know) which guarantees the evolved state satisfies the axiom defining states:

$$\|E(\hat{x};\Lambda_0(t))\hat{U}_{\Lambda(t)}(t)\psi\| = ?$$

– the Paley–Wiener theorem is again relevant.

It would seem that even in abandoning complete precision it is not possible to adopt the intuitively appealing localised time evolution axiom without losing what accuracy there might remain. Short of finding a way to make $t_1 > t_2$ (some limitation on the class of localised states, or on $\epsilon$ ), the only way to retain locality is to modify the time evolution of the localising function to explicitly guarantee that states evolve into states.

§§5.5 *In Conclusion*

The dogma behind the 'boundary of localisation' is that this is a place where the wave-function is insignificant. The localising isometry is employed to tidy this insignificant part into a finite volume that is as small as desired. Yet if this is truly an insignificant part of the state, why does it matter where it is? There is clearly no reason why the boundary of localisation cannot be as large as one cares to imagine, even extending to spatial infinity. Of course, this would require the abandonment of the present Principle of Locality, but, given the nature of this boundary, at an unmeasurable level.

Is the form adopted here of the Principle of Locality the right one? Should it be exact or only approximate? In the context of a Galilei–invariant theory there are none of the phenomena that make local theories so attractive; there is no speed of light limit. Indeed, the form of Hamiltonian mechanics from which Schrödinger's wave mechanics is derived is explicitly non-local. It is a trifle inconsistent to insist on a localised quantum theory that is constructed from an action–at–a–distance classical theory.

The most obvious way around this is to move to a Poincaré–invariant theory. This may be an obvious move, but as I have indicated in Chapter 4, there does not appear to be a credible Poincaré–invariant theory that but lacks locality. Indeed, of the two invariant volume elements on space-time, one ( $(d\sigma) = (dx^1 dx^2 dx^3,\ dx^0 dx^2 dx^3,\ dx^0 dx^1 dx^3,\ dx^0 dx^1 dx^2)$ ) produces invariant integrands of the form $j^\nu\, d\sigma_\nu$, but, since these are integrated over a time-like hypersurface that transforms with changes of coordinates, the integrals are specific to a preferred hypersurface (i.e., coordinate system); and the other ( $d^4x$ ) cannot sensibly be normalised. It is therefore not going to be possible to localise Lorentz boosts along with the rest of the Poincaré group in the manner presented here, since there is no 'global' unitary representation on space-time wave-functions. It is, of course, possible to use a hypersurface that is not invariant with respect to the entirety of the Poincaré group; one such approach will be examined in the next chapter; others have already been scrutinised in Chapter 4.

The localisation manufactured in this chapter is unsatisfactory for the loss of precision required, and then for the fact that this sacrifice is not enough to give a unique or appealing theory.

It is pertinent to mention that there are interpretations of the experimental realisations of the Einstein–Podolsky–Rosen thought-experiment that claim there is a non-local or action–at–a–distance effect.[9] To persist in constructing theories that are local by assumption, even in a weakened sense, is not, therefore, the indisputable way forward.

# §6 Bibliography

(References are preceded with the page number on which they first appear.)

2: R. Haag, D. Kastler, *An Algebraic Approach To Quantum Field Theory,* J. Math. Phys. **3** (1964) 848–861.
2: [1] D. R. E. Timson, *Locality in Non-Relativistic and Relativistic Quantum Mechanics,* Ph. D. thesis, University of St. Andrews, 1986. (And references therein.)
3: [2] D. R. E. Timson, K. K. Wan, '*Problems in Quantum Physics,*' Gdansk 1987 (World Scientific, Singapore 1988) ed. L. Kostro; pp 794–821.
cf. K. K. Wan, '*Local Observables, Nonlocality and Asymptotically Separable Quantum Mechanics,*' Found. Phys. (1988) **18** 887.
4: [3] *ibid.*
4: [4] cf. K. R. Popper, *Quantum Theory and the Schism In Physics,* Hutchinson 1982, pp 49, 53–64, 144–147.
9: [5] Nikolaĭ Nikolaevich Bogolûbov, Anatoliĭ Logunov, Ivan T. Todorov, *Introduction to Axiomatic Quantum Field Theory,* W. A. Benjamin 1975, pp 594–5.
10: [6] L. L. Foldy, *Relativistic Particle Systems With Interaction,* Phys. Rev. **122** (1961) 275–288.
10: [7] D. G. Currie, T. F. Jordan, E. C. G. Sudarshan, *Relativistic Invariance and Hamiltonian Theories Of Interacting Particles,* Rev. Mod. Phys. **35** (1963) 350–375.
12: [8] John F. Clauser, Micheal A. Horne, Abner Shimony, R. A. Holt, Phys. Rev. Lett. **23** (1969) 880.
15: [9] N. Cufaro-Petroni, C. Dewdney, P. R. Holland, A. Kyprianidis, J.-P. Vigier, *EPR constraints on quantum action–at–a–distance: the Sutherland paradox,* Found. Phys. **17** (1987) 759–773.

# Chapter 7

## The Light-cone Mechanics of G. H. Derrick

*a critical review*

*'A man should never be ashamed to own he has been in the wrong, which is saying, in other words, that he is wiser today than he was yesterday.'*

Jonathan Swift **On Various Subjects.**

## §1 A Critiqué

I am not going to attempt any préçis of Graham Derrick's papers[1] for two reasons: I do not think any summary of mine could add anything to the published account, and the revision of Derrick's ideas that occurs below will recapitulate large portions of his work in only a mildly different form. This is not to say that I find Derrick's idea to be especially good.

In 1949 Dirac proposed a number of Hamiltonian formulations for a Poincaré–relativistic mechanics[2]; each was based on a different constraint (hence the subject spawned by this paper is called *Constraint Hamiltonian Dynamics*) and thus on a different sub-group of the continuous Poincaré transformations. One of these, apparently less explored than the others, was based on the light-cone. Dirac argued that, since the light-cone is not invariant under 4–translations, just as Galilean space is not invariant under time translations, so there must be four Hamiltonians in this system of dynamics: the momentum 4–vector. This idea has considerable appeal; the alternative is, most commonly, to use constant-time hyperplanes, which are not invariant under Lorentz boosts and time translations – surely a less symmetrically balanced situation?

Derrick takes on Dirac's idea of using a backwards–in–time light-cone but then makes two errors: an error of interpretation and an error of formulation. The theory that follows is by no means fatally affected by these errors, and only a modest re-writing is required to improve matters.

The error of interpretation concerns the nature of initial-value problems, and how the process of their solution is to be viewed in the realms of physics. It is, naturally, unavoidable that the mathematical expression of a theory be linked to the physical reality it is to describe. However, not all of the mathematics and mathematical processes necessarily have a physical counterpart. It is especially important to note this when it appears that a simple, physical interpretation of a mathematical process for one theory does not have a correspondence in a second, albeit similar, theory. Thus in Galilei-relativistic mechanics, and certainly in Schrödinger's wave mechanics, the process of solving the initial-value problem is called 'the evolution of the system'; the initial value is the state of the system at some instant and the solution of the problem for subsequent instants is, likewise, called the instantaneous state of the system.

Poincaré–relativistic mechanics is obviously similar to Galilei–relativistic mechanics; it does not follow from this, however, that the Galilean interpretation of the initial-value problem is correct for a Poincaré–relativistic theory. It may be, but it is not obvious *a priori*. Derrick's discussion of observers, the quality of their information, and various hypersurfaces in space-time is, therefore, at least premature.

In my opinion, much of Derrick's interpretational preamble is debatable. Just as constant-time hyperplanes in Galilean theories provide that all influences (forces and the like) act instantaneously everywhere, and so propagate with infinite velocity; so, by taking backwards–in–time light-cones, Derrick now propounds a theory in which all influences surely propagate at the speed of light. This must limit the applicability of the theory of mechanics being developed. Also, it is all very well to say, vaguely, that the only things causally influencing the value on the light-cone are the events within it; it is a bit more challenging to show that such causes occur within the bounds of the Special Principle of Relativity.

By insisting on an observer-centred analysis, Derrick does not make the case for his theory any stronger. He suggests the introduction of a limiting hypothesis: nothing beyond some particular light-cone (the initial one, I suppose) is important. This, it is admitted, is not particularly useful nor comprehensive. The suggestion that there may be insufficient initial data to determine the motion of the system does not inspire confidence. These are not, in fact, real problems. In the first place, there are no surprises in mechanics; it is in this sense that mechanics is deterministic: given an initial value, the value of all fields throughout space-time, and the evolution that this implies (i.e., a complete specification of the problem), then there is a unique solution. This solution is the complete motion of the system through space-time, from the initial value to what might be termed 'evolutionary infinity' – in Galilean theories this is temporal infinity. The solution of a problem in mechanics is primarily a description of the whole system, and only as such, as a view of all space-time, and only as such as a 'prediction'. The idea that an observer predicts the values on his backward light-cone as he evolves with it is no different than Laplace's demon in Galilean mechanics, and is just as unsupportable[3].

The second error in Derrick's thesis is an error of formulation. In contrast to Dirac's musings, Derrick proposes only one Hamiltonian, though it is parameterised by the 4–velocity of the observer, and hence the proper time, $\tau$, of the observer. At first blush there seems nothing very wrong with

this. Upon reflection, this is not the case because:

(i) the frame of reference in which the mechanical problem is to be analysed may be taken to be the rest frame of the observer – and, arguably, this must be so –;

(ii) if $\tau$ is the proper time of the observer then it is generally not the proper time for the system being observed, the only interesting 4–velocity that arises from $\tau$ is, therefore, the 4–velocity of the observer.

The difficulty here is that, unlike the universal time of Galilean mechanics, a proper time is neither unique nor very widely relevant: it is highly local in nature and inevitably tied to a specific world-line. Indeed, the introduction of 'the observer's 4–velocity' begs the question: in whose coordinates is the formalism written?

(a) The observer's, whence the evolution is in terms of $x^0$ and the 4–velocity is, simply, $(\mathbf{1}, \mathbf{0})$.

(b) The observed system, in which case it can only be point-like – to avoid the pitfalls catalogued in chapter 4, §2, and so give an unambiguous meaning to $\tau$ – and will have a rather trivial world-line: the definition of the observer's 4–velocity is no longer correct; and the problem is now to determine the motion of the observer, a strange rôle reversal.

(c) Some other observer's rest frame, in which case why not abandon the first observer in favour of this one?

Once it is accepted that the frame of reference in which the problem is posed is the rest frame of the observer there are two consequences. The effect of a Lorentz boost is to change not only the coordinate system, but also the observer. The second consequence arises from this: the parameter $\tau$ is not an invariant but is always to be taken as the time coordinate of the current frame of reference. The use of the observer as an aid to interpreting this theory is now at an end: the one useful function that may be served by an observer is to be at rest with respect to a coordinate system, as is conventionally the case in classical mechanics. To persist in phrasing an interpretation in terms of observations is to insist on clumsy contrivances – such as having 'auxiliary observers' throughout space-time sending coded signals to the 'main observer' – that make the whole idea seem foolish.

Having dispensed with the notion that a proper time can be used as the (single) evolution parameter, I am left with the time coordinate as such a parameter. Since the time coordinate is only one component of a 4–vector, to establish a Poincaré–invariant theory it therefore follows that there are four Hamiltonians at work – just as Dirac surmised.

## §2 A Revision

Consider the whole approach afresh. Take a reference frame $X$ with coordinates $(x)$, and select an arbitrary event that will be denoted by its coordinates in $X$, viz. $(z)$. The backward light-cone at $(z)$ can be parameterised by the spatial coordinates, $(\mathbf{y})$, of the point on this hypersurface, the time (or time-like) coordinate, $y^0$, is then $-|\mathbf{y}|$.

The events of Minkowski space-time can now be expressed in terms of the usual coordinates of $X$, or by way of the 3–vector $\mathbf{y}$ and any one of the components of the 4–vector $(z)$, which will be allowed to vary, or *evolve* parallel to the corresponding $x^\nu$–axis. These two coordinate systems are related by the four equations

$$x^\nu = z^\nu + y^\nu \tag{2.1}$$

or,

$$\left.\begin{aligned} x^0 &= z^0 - |\mathbf{y}| \\ \mathbf{x} &= \mathbf{z} + \mathbf{y} \end{aligned}\right\} \tag{2.2}$$

The metric for the coordinate system $(x)$ is the usual one: to obtain the metric for what will now be called *the light-cone coordinate system*, the equations 1 or 2 are regarded as a coordinate transformation. Denote the metric of $(x)$ by

$$\eta = \mathrm{diag}[+1, -1, -1, -1] \quad ,$$

and the metric of the new coordinates by $\overline{\eta}$. Clearly, there are four light-cone coordinate systems depending on which of the four components of $(z)$ is chosen as the evolution parameter. There are,

as might be expected, only two cases to consider: a time-like evolution parameter or a space-like one. To distinguish the coordinates of the apex of the light-cone from the light-cone coordinate system, denote the latter (the transformed coordinate system) by sets of four coordinates $(\overline{x})$ – note that $\overline{x}$ is not a 4–vector, even if it is written in the form of one as a convenience.

§§2.1 *Case 0:* $\overline{x}^0 = x^0$

Equation 2 now reads

$$\begin{aligned} x^0 &= \overline{x}^0 - |\overline{\mathbf{x}}| \\ \mathbf{x} &= \mathbf{z} + \overline{\mathbf{x}} \end{aligned} \tag{2.3}$$

with inverse

$$\begin{aligned} \overline{x}^0 &= x^0 + |\mathbf{x} - \mathbf{z}| \\ \overline{\mathbf{x}} &= \mathbf{x} - \mathbf{z} \end{aligned} \tag{2.4}$$

The Jacobian of this transformation is 1.

Now

$$\begin{aligned} \overline{\eta}_{\mu\nu} &= \eta_{\rho\sigma} \frac{\partial x^\rho}{\partial \overline{x}^\mu} \frac{\partial x^\sigma}{\partial \overline{x}^\nu} \\ &= \frac{\partial x^0}{\partial \overline{x}^\mu} \frac{\partial x^0}{\partial \overline{x}^\nu} - \sum_{l=1}^{3} \frac{\partial x^l}{\partial \overline{x}^\mu} \frac{\partial x^l}{\partial \overline{x}^\nu} \end{aligned}$$

whence

$$\overline{\eta} = \begin{pmatrix} 1 & -\frac{\overline{x}^1}{|\overline{\mathbf{x}}|} & -\frac{\overline{x}^2}{|\overline{\mathbf{x}}|} & -\frac{\overline{x}^3}{|\overline{\mathbf{x}}|} \\ -\frac{\overline{x}^1}{|\overline{\mathbf{x}}|} & \left(\frac{\overline{x}^1}{|\overline{\mathbf{x}}|}\right)^2 - 1 & \frac{\overline{x}^1 \overline{x}^2}{|\overline{\mathbf{x}}|^2} & \frac{\overline{x}^1 \overline{x}^3}{|\overline{\mathbf{x}}|^2} \\ -\frac{\overline{x}^2}{|\overline{\mathbf{x}}|} & \frac{\overline{x}^1 \overline{x}^2}{|\overline{\mathbf{x}}|^2} & \left(\frac{\overline{x}^2}{|\overline{\mathbf{x}}|}\right)^2 - 1 & \frac{\overline{x}^2 \overline{x}^3}{|\overline{\mathbf{x}}|^2} \\ -\frac{\overline{x}^3}{|\overline{\mathbf{x}}|} & \frac{\overline{x}^1 \overline{x}^3}{|\overline{\mathbf{x}}|^2} & \frac{\overline{x}^2 \overline{x}^3}{|\overline{\mathbf{x}}|^2} & \left(\frac{\overline{x}^3}{|\overline{\mathbf{x}}|}\right)^2 - 1 \end{pmatrix} \tag{2.5}$$

This metric has a singularity at $\overline{\mathbf{x}} = 0$ that will need to be considered further.

The action integral is taken to be

$$\int L_{(0)}\, d\overline{x}^0 = -mc \int \sqrt{\overline{\eta}_{\mu\nu} \frac{d\overline{x}^\mu}{d\overline{x}^0} \frac{d\overline{x}^\nu}{d\overline{x}^0}}\, d\overline{x}^0 \tag{2.6}$$

So the Lagrangian is

$$\begin{aligned} L_{(0)} &= -mc \sqrt{1 - 2\sum_{l=1}^{3} \frac{\overline{x}^l}{|\overline{\mathbf{x}}|} \frac{d\overline{x}^l}{d\overline{x}^0} + \sum_{l=1}^{3} \left( \left(\frac{\overline{x}^l}{|\overline{\mathbf{x}}|}\right)^2 - 1 \right) \left(\frac{d\overline{x}^l}{d\overline{x}^0}\right)^2 + 2 \sum_{\substack{k,l=1 \\ k \neq l}}^{3} \frac{\overline{x}^l \overline{x}^k}{|\overline{\mathbf{x}}|^2} \frac{d\overline{x}^l}{d\overline{x}^0} \frac{d\overline{x}^k}{d\overline{x}^0}} \\ L_{(0)} &= -mc \sqrt{\left(1 - \sum_{l=1}^{3} \frac{\overline{x}^l}{|\overline{\mathbf{x}}|} \frac{d\overline{x}^l}{d\overline{x}^0}\right)^2 - \sum_{l=1}^{3} \left(\frac{d\overline{x}^l}{d\overline{x}^0}\right)^2} \end{aligned} \tag{2.7}$$

The conjugate momenta are then

$$\begin{aligned} \pi^l &= \frac{\partial L_{(0)}}{\partial \left(\frac{d\overline{x}^l}{d\overline{x}^0}\right)} \\ &= -\frac{m^2 c^2}{L_{(0)}} \left( \frac{\overline{x}^l}{|\overline{\mathbf{x}}|} \left( 1 - \sum_{k=1}^{3} \frac{\overline{x}^k}{|\overline{\mathbf{x}}|} \frac{d\overline{x}^k}{d\overline{x}^0} \right) + \frac{d\overline{x}^l}{d\overline{x}^0} \right) \end{aligned} \tag{2.8}$$

The Hamiltonian is, by the usual Legendre transformation,

$$\begin{aligned} H_{(0)} &= \sum_{l=1}^{3} \pi^l \frac{d\overline{x}^l}{d\overline{x}^0} - L_{(0)} \\ &= \frac{m^2c^2\left(1 - \sum_{l=1}^{3} \frac{\overline{x}^l}{|\overline{\mathbf{x}}|}\frac{d\overline{x}^l}{d\overline{x}^0}\right)}{-L_{(0)}} \end{aligned} \tag{2.9}$$

It is then a straightforward matter to show what is a special case in Derrick's analysis, though here it will be accorded greater merit:

$$H_{(0)} = \frac{|\overline{\mathbf{x}}|}{2}\boldsymbol{\pi}\cdot\boldsymbol{\pi} + \frac{m^2c^2}{\overline{\mathbf{x}}\cdot\boldsymbol{\pi}} \tag{2.10}$$

§§2.2 *Case i:* $\overline{x}^0 = z^i$

Take $\overline{x}^0 = z^1$. Equation 2 now reads

$$\begin{aligned} x^0 &= z^0 - |\overline{\mathbf{x}}| \\ x^1 &= \overline{x}^0 + \overline{x}^1 \\ x^j &= z^j + \overline{x}^j \qquad (j = 2,3) \end{aligned}$$

with inverse

$$\left.\begin{aligned} \overline{x}^0 &= x^1 \mp \sqrt{(x^0 - z^0)^2 - (x^2 - z^2)^2 - (x^3 - z^3)^2} \\ \overline{x}^1 &= \pm\sqrt{(x^0 - z^0)^2 - (x^2 - z^2)^2 - (x^3 - z^3)^2} \\ \overline{x}^j &= z^j + \overline{x}^j \qquad (j = 2,3) \end{aligned}\right\} \tag{2.11}$$

The Jacobian of this transformation is

$$J = \frac{\overline{x}^1}{|\overline{\mathbf{x}}|}$$

which has a singularity at the origin and vanishes when $\overline{x}^1 = 0$.

The process of generating momenta and a Hamiltonian can then be carried out as normal.

$$\overline{\eta}_{(1)} = \begin{pmatrix} -1 & -1 & 0 & 0 \\ -1 & \left(\frac{\overline{x}^1}{|\overline{\mathbf{x}}|}\right)^2 - 1 & \frac{\overline{x}^1\overline{x}^2}{|\overline{\mathbf{x}}|^2} & \frac{\overline{x}^1\overline{x}^3}{|\overline{\mathbf{x}}|^2} \\ 0 & \frac{\overline{x}^1\overline{x}^2}{|\overline{\mathbf{x}}|^2} & \left(\frac{\overline{x}^2}{|\overline{\mathbf{x}}|}\right)^2 - 1 & \frac{\overline{x}^2\overline{x}^3}{|\overline{\mathbf{x}}|^2} \\ 0 & \frac{\overline{x}^1\overline{x}^3}{|\overline{\mathbf{x}}|^2} & \frac{\overline{x}^2\overline{x}^3}{|\overline{\mathbf{x}}|^2} & \left(\frac{\overline{x}^3}{|\overline{\mathbf{x}}|}\right)^2 - 1 \end{pmatrix} \tag{2.12}$$

Whence the Lagrangian, $L_{(1)}$, may be obtained:

$$L_{(1)} = -mc\sqrt{-\left(1 + \frac{d\overline{x}^1}{d\overline{x}^0}\right)^2 - \left(\frac{d\overline{x}^2}{d\overline{x}^0}\right)^2 - \left(\frac{d\overline{x}^3}{d\overline{x}^0}\right)^2 + \left(\frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\right)^2} \tag{2.13}$$

The Hamiltonian is then

$$H_{(1)} = \frac{m^2c^2\left(1 + \frac{d\overline{x}^1}{d\overline{x}^0}\right)}{L_{(1)}} \tag{2.14}$$

## §3 A Generalisation

The choice of light-cones as the hypersurfaces on which to base a new analysis is not, as the introductory critiqué mentions, uniquely obvious. Any hypersurface invariant under a (continuous) sub-group of the Poincaré group would do about as well. The reasons Derrick gives for his choice are not completely satisfactory. Very much the same level of satisfaction (and no singularities) is to be had by using backwards–in–time hyperboloids. If the coordinate transformation (equation 2.2) is written in the general form ($z^0$ being the evolution parameter again):

$$\left.\begin{aligned} x^0 &= \overline{x}^0 + \overline{\overline{x}}^0 \\ \mathbf{x} &= \mathbf{z} + \overline{\mathbf{x}} \end{aligned}\right\} \tag{3.1}$$

where

$$z^0 = \overline{\overline{x}}^0 = \overline{\overline{x}}^0(\overline{\mathbf{x}}, b)$$

then I can actually deal with any space-like hypersurface for most of the analysis of the classical problem. The additional parameter, $b$, is defined to be an invariant that in some way characterises the surface – clearly not something of the most general form, but for the light-cone and the hyperboloids it has an immediate rôle in that it allows all of them to be considered together: for these surfaces

$$\overline{\overline{x}}^0 = -\sqrt{\overline{\mathbf{x}} \cdot \overline{\mathbf{x}} + b^2} \tag{3.2}$$

and the limit $b \to 0$ recovers the light-cone.

Now

$$\frac{\partial x^0}{\partial \overline{x}^0} = 1 \qquad \frac{\partial x^0}{\partial \overline{x}^l} = \frac{\partial \overline{\overline{x}}^0}{\partial \overline{x}^l} = \xi_l$$

$$\frac{\partial x^l}{\partial \overline{x}^0} = 0 \qquad \frac{\partial x^l}{\partial \overline{x}^j} = \delta^l_j$$

This gives a well-behaved Jacobian

$$J = \begin{vmatrix} 1 & \xi_1 & \xi_2 & \xi_3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1$$

The metric tensor transforms to

$$\overline{\eta}_{(0)} = \begin{pmatrix} 1 & \xi_1 & \xi_2 & \xi_3 \\ \xi_1 & (\xi_1)^2 - 1 & \xi_1\xi_2 & \xi_1\xi_3 \\ \xi_2 & \xi_1\xi_2 & (\xi_2)^2 - 1 & \xi_2\xi_3 \\ \xi_3 & \xi_1\xi_3 & \xi_2\xi_3 & (\xi_3)^2 - 1 \end{pmatrix} \tag{3.3}$$

The Lagrangian is then (using the notation established above, together with $\overline{u}^i = d\overline{x}^i / d\overline{x}^0$ )

$$\begin{aligned} L_{(0)} &= -mc\sqrt{\overline{\eta}_{(0)\mu\nu} \frac{d\overline{x}^\mu}{d\overline{x}^0} \frac{d\overline{x}^\nu}{d\overline{x}^0}} \\ &= -mc\sqrt{1 + 2\sum_l \xi_l \overline{u}^l + 2\xi_1\xi_2\overline{u}^1\overline{u}^2 + 2\xi_2\xi_3\overline{u}^2\overline{u}^3 + 2\xi_1\xi_3\overline{u}^1\overline{u}^3 + \sum_l \left((\xi_l)^2 - 1\right)(\overline{u}^l)^2} \\ &= -mc\sqrt{1 + 2\boldsymbol{\xi}\cdot\overline{\mathbf{u}} + (\boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 - \overline{\mathbf{u}}\cdot\overline{\mathbf{u}}} \\ &= -mc\sqrt{(1 + \boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 - \overline{\mathbf{u}}\cdot\overline{\mathbf{u}}} \end{aligned} \tag{3.4}$$

The conjugate momenta and the Hamiltonian follow readily enough:

$$\begin{aligned} \pi_l &= \frac{\partial L_{(0)}}{\partial \overline{u}^l} \\ &= \frac{m^2c^2}{L_{(0)}} \left(\xi_l(1 + \boldsymbol{\xi}\cdot\overline{\mathbf{u}}) - \overline{u}^l\right) \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} H_{(0)} &= \boldsymbol{\pi}\cdot\overline{\mathbf{u}} - L_{(0)} \\ &= -\frac{m^2c^2}{L_{(0)}}(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}}) \end{aligned} \tag{3.6}$$

Quantities proportional to $(m^2c^2/L_{(0)})^2$ are

$$\begin{aligned} H_{(0)}^2 \quad &: \quad (1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 \\ m^2c^2 \quad &: \quad (1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 - \overline{\mathbf{u}}\cdot\overline{\mathbf{u}} \end{aligned}$$

i.e.,

$$\begin{aligned} H_{(0)}^2 - m^2c^2 \quad &: \quad \overline{\mathbf{u}}\cdot\overline{\mathbf{u}} \\ \boldsymbol{\pi}\cdot\boldsymbol{\pi} \quad &: \quad \boldsymbol{\xi}\cdot\boldsymbol{\xi}(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 - 2\boldsymbol{\xi}\cdot\overline{\mathbf{u}}(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}}) + \overline{\mathbf{u}}\cdot\overline{\mathbf{u}} \\ \sum_l \pi_l H_{(0)} \quad &: \quad -\sum_l \xi_l(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 + \sum_l \overline{u}^l(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}}) \end{aligned}$$

i.e.,

$$\boldsymbol{\xi}\cdot\boldsymbol{\pi}H_{(0)} \quad : \quad -\boldsymbol{\xi}\cdot\boldsymbol{\xi}(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}})^2 + \boldsymbol{\xi}\cdot\overline{\mathbf{u}}(1+\boldsymbol{\xi}\cdot\overline{\mathbf{u}})$$

So

$$\begin{aligned} \boldsymbol{\pi}\cdot\boldsymbol{\pi} &= -2\boldsymbol{\xi}\cdot\boldsymbol{\pi}H_{(0)} - \boldsymbol{\xi}\cdot\boldsymbol{\xi}H_{(0)}^2 + H_{(0)}^2 - m^2c^2 \\ \boldsymbol{\pi}\cdot\boldsymbol{\pi} + m^2c^2 &= (1-\boldsymbol{\xi}\cdot\boldsymbol{\xi})H_{(0)}^2 - 2\boldsymbol{\xi}\cdot\boldsymbol{\pi}H_{(0)} \\ &= (1-\boldsymbol{\xi}\cdot\boldsymbol{\xi})\left(H_{(0)}^2 - \frac{2(\boldsymbol{\xi}\cdot\boldsymbol{\pi})H_{(0)}}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}} - \left(\frac{\boldsymbol{\xi}\cdot\boldsymbol{\pi}}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}\right)^2 + \left(\frac{\boldsymbol{\xi}\cdot\boldsymbol{\pi}}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}\right)^2\right) \\ &= (1-\boldsymbol{\xi}\cdot\boldsymbol{\xi})\left(H_{(0)} - \frac{\boldsymbol{\xi}\cdot\boldsymbol{\pi}}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}\right)^2 - \left(\frac{\boldsymbol{\xi}\cdot\boldsymbol{\pi}}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}\right)^2 \end{aligned}$$

Whence, taking the positive square-root:

$$H_{(0)} = \frac{1}{\sqrt{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}}\sqrt{\boldsymbol{\pi}\cdot\boldsymbol{\pi} + m^2c^2 + \frac{(\boldsymbol{\xi}\cdot\boldsymbol{\pi})^2}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}} + \frac{\boldsymbol{\xi}\cdot\boldsymbol{\pi}}{1-\boldsymbol{\xi}\cdot\boldsymbol{\xi}} \tag{3.7}$$

For the hyperboloid–light-cone case

$$\boldsymbol{\xi} = -\frac{\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}\cdot\overline{\mathbf{x}} + b^2}}$$

So that

$$H_{(0)} = \frac{\sqrt{\overline{\mathbf{x}}\cdot\overline{\mathbf{x}} + b^2}}{|b|}\sqrt{\boldsymbol{\pi}\cdot\boldsymbol{\pi} + m^2c^2 + \frac{(\overline{\mathbf{x}}\cdot\boldsymbol{\pi})^2}{b^2}} - \frac{\sqrt{\overline{\mathbf{x}}\cdot\overline{\mathbf{x}} + b^2}}{b^2}\overline{\mathbf{x}}\cdot\boldsymbol{\pi}$$

– the quantisation of this would require a satisfactory solution to the problem of quantising non-commuting observables, which is not something I have seen yet. The strategy adopted by Derrick must correspond to the choice of one of the numerous quantisation rules; there is no indication that this is any better or worse than rival rules.

## §4 A Worked Example

There is one instance in which a quantisation process appears to exist. This is the case of a light-cone in $(1+1)$–dimensions. Having said that, hardly any aspect of the resultant quantum theory is without problems.

From equations 2.7 and 2.8

$$\pi = \frac{mc\left(\frac{\overline{x}}{|\overline{x}|}\left(1-\frac{\overline{x}}{|\overline{x}|}\frac{d\overline{x}}{d\overline{x}^0}\right)+\frac{d\overline{x}}{d\overline{x}^0}\right)}{\sqrt{1-2\frac{\overline{x}}{|\overline{x}|}\frac{d\overline{x}}{d\overline{x}^0}}}$$

i.e.,

$$\pi = \frac{mc\frac{\overline{x}}{|\overline{x}|}}{\sqrt{1-2\frac{\overline{x}}{|\overline{x}|}\frac{d\overline{x}}{d\overline{x}^0}}}$$

And

$$H_{(0)} = \frac{\pi^2+m^2c^2}{2\frac{\overline{x}}{|\overline{x}|}\pi}$$

Consider now the two parts of the light-cone:

$\overline{x}>0$ : $\quad \pi = \dfrac{mc}{\sqrt{1-2\frac{d\overline{x}}{d\overline{x}^0}}} > 0 \qquad H_{(0)} = \dfrac{\pi}{2}+\dfrac{m^2c^2}{2\pi} > 0$

$\overline{x}<0$ : $\quad \pi = \dfrac{-mc}{\sqrt{1+2\frac{d\overline{x}}{d\overline{x}^0}}} < 0 \qquad H_{(0)} = -\dfrac{\pi}{2}-\dfrac{m^2c^2}{2\pi} > 0$

By a straightforward calculation (to be discussed in detail in §5 below),

$$\frac{d\overline{x}}{d\overline{x}^0} \in \left(-\infty, \frac{1}{2}\right)$$

so that $|\pi| \in (0,\infty)$.

To quantise this there is no doubt which Hilbert space to choose:

$$\mathcal{H} = L^2\left(\Re, \frac{d\overline{x}}{\sqrt{|\overline{x}|}}\right)$$

It now becomes clear why Derrick chose to introduce anti-particles at this stage: he wanted to use

$$\hat{\pi} = -\sqrt{|\overline{x}|}i\hbar\frac{\partial}{\partial\overline{x}}\frac{1}{\sqrt{|\overline{x}|}} \tag{4.1}$$

as a momentum observable, but this has spectrum $(-\infty,\infty)$, whereas, if $\overline{x}>0$, it is apparent that the spectrum should be $(0,\infty)$. Now, in a curious inversion of the situation Stückelberg found himself in, Derrick postulates that when a particle has momentum, $\pi$, a corresponding anti-particle will have momentum, $-\pi$. This may be obtained by taking the negation of the Lagrangian and leads to the negation of the Hamiltonian, which is now always negative. This is by no means a straightforward application of what is commonly called the *Feynman–Stückelberg interpretation*, but is not outrageously different. A more rigorous application of Feynman and Stückelberg's idea would involve the use of forwards–in–time light-cones (or hyperboloids) for anti-particles, since the basis for the interpretation is time symmetry.

Actually, formula 1 will not do. If the right-hand side is the quantisation of anything then it must be

$$\frac{\overline{x}}{|\overline{x}|}\pi \quad .$$

If this is not so then the momentum space representation will freely mix particle and anti-particle components: any $\tilde{\phi}(\pi)$, for positive $\pi$, will contribute to the state of the particle if $\overline{x}>0$, and to that of the anti-particle for $\overline{x}<0$ ! Time evolution cannot be formulated in such chaos. I am forced

to conclude that Derrick has not thought through some of the statements he makes in his second paper (his §2).

Assume now that

$$\hat{p} = \widehat{\frac{\overline{x}}{|\overline{x}|}\pi} = -\sqrt{|\overline{x}|}i\hbar\frac{\partial}{\partial\overline{x}}\frac{1}{\sqrt{|\overline{x}|}} \tag{4.2}$$

This is quite similar to the quantity Derrick denotes by $D$ in terms of which he invokes the Feynman–Stückelberg interpretation. Here, however, instead of using two copies of a momentum space, the spectrum of $\hat{p}$ is split: $p > 0$ for particles, and $p < 0$ for anti-particles. Also

$$\hat{H} = \frac{(\hat{p})^2 + m^2c^2}{2\hat{p}} \tag{4.3}$$

– a formula valid for particle and anti-particle alike.

Having apparently managed to quantise one example, I can snatch defeat from the jaws of victory by making the two observations:

(i) no state involving only a particle can be confined to a bounded subset of the light-cone;

(ii) no state that is initially confined to a bounded subset of the light-cone will remain so for more than that instant (a single value of $z^0$).

Both of these statements are proven below using the Paley–Wiener theorem.

Configuration space and momentum space are connected by a form of Fourier transform:

$$\tilde{\phi}(p) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} d\overline{x}\, e^{ip\overline{x}/\hbar}\,\phi(\overline{x},0) \tag{4.4}$$

$$\phi(\overline{x},t) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dp\, e^{-ip\overline{x}/\hbar}U_t\tilde{\phi}(p) \tag{4.5}$$

(i) For a state that involves no anti-particle content,

$$\text{supp}\,\tilde{\phi}(p) \subset (0,\infty)$$

But this cannot be the restriction (to a real argument) of an analytic function because of the identity theorem[4], and so, by the contra-positive of the Paley–Wiener theorem, $\phi(\overline{x})$ cannot have compact support.

(ii) If $\phi(\overline{x},0)$ is of compact support (as a function of $\overline{x}$) then $\tilde{\phi}(p)$ will be an exponential entire function (Paley–Wiener theorem). However, time evolution is generated by

$$U_t = \exp\left(\frac{it}{\hbar}\frac{p^2+m^2c^2}{2p}\right)$$

and the exponent has a simple pole at $p = 0$, therefore

$$\tilde{\phi}(p)\exp\left(\frac{it}{\hbar}\frac{p^2+m^2c^2}{2p}\right)$$

is not analytic. Once again, the contra-positive of the Paley–Wiener theorem says that $\phi(\overline{x},t)$ will not be of compact support if $t \neq 0$.

This whole exercise is probably best treated as an illustration of how easy it is to produce a mathematical formalism that is like a quantum theory while being nothing of the sort. This is not a conclusion I draw solely from the existence of a *No Go* theorem in the style of Hegerfeldt, unhelpful as this may be. The symmetry properties of equation 5 are, also, extremely dubious. Further, there is no mention of the Hamiltonian, $H_1$ , and the momentum variable by way of which it is defined.

## §5 A Consistency Check

### §§5.1 *Velocities, real and imagined*

It is tempting to call the three-component object

$$\overline{\mathbf{u}} = \left(\frac{d\overline{x}^l}{d\overline{x}^0}\right)$$

the velocity of the system under observation. There is a connection with this physical concept but it is by no means clear-cut.

If a trajectory is chosen,

$$\begin{aligned}\mathbf{x} &= \mathbf{x}(\overline{x}^0)\\ &= \overline{\mathbf{x}}(\overline{x}^0) + \mathbf{z}\\ &= \overline{\mathbf{x}}(x^0 + |\mathbf{x}-\mathbf{z}|) + \mathbf{z}\end{aligned}$$

Then, differentiating this formula,

$$\frac{d\mathbf{x}}{da} = \frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\left(\frac{dx^0}{da} + \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\frac{d(\mathbf{x}-\mathbf{z})}{da}\right) + \frac{d\mathbf{z}}{da}$$

Whence, selecting $a = x^0$, and noting that $\mathbf{z}$ is independent of $x^0$ – the observer is not moving, if you like:

$$\frac{d\mathbf{x}}{d\mathbf{x}^0} = \frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\left(1 + \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\frac{d\mathbf{x}}{dx^0}\right)$$

and therefore

$$\frac{d\overline{\mathbf{x}}}{d\overline{x}^0} = \frac{\dfrac{d\mathbf{x}}{dx^0}}{1 + \dfrac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\dfrac{d\mathbf{x}}{dx^0}} \tag{5.1}$$

Now $dx^l/dx^0$ is, for want of a better expression, the actual velocity of the system. As such it obeys the Special Principle:

$$\left|\frac{dx^l}{dx^0}\right| < 1$$

Therefore, if the system is headed towards the observer at something close to the speed of light, then $d\overline{x}^l/d\overline{x}^0 \to \pm\infty$. A bizarre sort of velocity.

Note that the contravariant vector for velocity, an object that must be at least closer to the physical reality modelled, is written

$$u^\nu = \frac{\dfrac{dx^\nu}{dx^0}}{\sqrt{1 - \dfrac{d\mathbf{x}}{dx^0}\cdot\dfrac{d\mathbf{x}}{dx^0}}} \tag{5.2}$$

So that

$$\begin{aligned}\frac{d\overline{x}^l}{d\overline{x}^0} &= \frac{u^l}{\left(\sqrt{1 - \dfrac{d\mathbf{x}}{dx^0}\cdot\dfrac{d\mathbf{x}}{dx^0}}\right)^{-1} + \dfrac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\mathbf{u}}\\ &= |\mathbf{x}-\mathbf{z}|\frac{u^l}{u^0|\mathbf{x}-\mathbf{z}| + (\mathbf{x}-\mathbf{z})\cdot\mathbf{u}}\end{aligned}$$

But $(y) = (-|\mathbf{x}-\mathbf{z}|, \mathbf{x}-\mathbf{z})$, so

$$\frac{d\overline{x}^l}{d\overline{x}^0} = \frac{y^0 u^l}{u^\nu y_\nu} \quad . \tag{5.3}$$

a formula with somewhat involved symmetry characteristics.

The use of $\overline{\mathbf{u}}$ does make the inclusion of an observer a necessity in interpreting the theory so formulated, though what $\overline{\mathbf{u}}$ represents is a far from trivial question. From the nature of the light-cone coordinates, $\overline{\mathbf{u}}$ may reasonably be called the velocity that the observer (who was singled out from his equivalent brethren at the beginning of the formulation) 'sees' the system to have. $\overline{\mathbf{u}}$ is more an apparent velocity than a real one. Now an 'apparent quantity' may make some sense if it is something directly observed in an experiment. As a trial requirement, it may be posited that either space-time is uniformly illuminated, or every particle is continually giving off light signals in order that any observer can make his observations. This will allow the determination of that component of $\overline{\mathbf{u}}$ perpendicular to the light signals connecting particle and observer. There must also be some way for the last component to be found – for instance, if there is a way of determining how far away the particle is, i.e., how far along his line of sight the light signals have travelled. It is for this reason that Derrick devised his array of 'auxiliary observers' spread throughout space, which is as sensible an approach as any.

This approach to mechanics clearly involves considerably more mental baggage to make any sense. What makes $\overline{\mathbf{u}}$ an unphysical quantity is the absurdity that must be introduced into the interpretational scheme, and hence any experimental determination, to make it measurable. No-one would construct even a part of Derrick's set of auxiliary observers, especially when the actual velocity is easier to determine. A more direct problem is that this 'velocity' seems to be only four components of a second order tensor. This is an almost inevitable consequence of taking the ratio of two infinitesimals from unrelated 4–vectors. Of course, the interpretational imbroglio is only exacerbated in the context of a quantum theory.

§§5.2 *The Lagrangian for* $\overline{x}^0 = z^0$

The previous analysis is, it is to be imagined, not very different from any other formulation of the classical problem: the answer obtained should be essentially the same. To see this, consider my simplification of Derrick's Lagrangian again:

$$L_{(0)} = -mc\sqrt{\left(1 - \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\right)^2 - \frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\cdot\frac{d\overline{\mathbf{x}}}{d\overline{x}^0}}$$

using equation 1

$$= -mc\sqrt{\left(\frac{1}{1+\dfrac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\dfrac{d\mathbf{x}}{dx^0}}\right)^2 - \frac{\dfrac{d\mathbf{x}}{dx^0}\cdot\dfrac{d\mathbf{x}}{dx^0}}{\left(1+\dfrac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\dfrac{d\mathbf{x}}{dx^0}\right)^2}}$$

$$= -mc\frac{\sqrt{1 - \dfrac{d\mathbf{x}}{dx^0}\cdot\dfrac{d\mathbf{x}}{dx^0}}}{\left|1+\dfrac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\dfrac{d\mathbf{x}}{dx^0}\right|} \tag{5.4}$$

$$= mc\frac{y^0\sqrt{\eta_{\mu\nu}\dfrac{dx^\mu}{dx^0}\dfrac{dx^\nu}{dx^0}}}{\left|\dfrac{dx^\nu}{dx^0}y_\nu\right|} \tag{5.5}$$

This is a well-defined quantity that has a clear, if not particularly simple, relation to the usual Lagrangian.

§§5.3 *Singularities and the Hamiltonian*

It was noted in passing that there was a singular point in the metric for the light-cone coordinates. The formula for the Hamiltonian, equation 2.10 above,

$$H_{(0)} = \frac{|\overline{\mathbf{x}}|}{2}\frac{\pi\cdot\pi + m^2c^2}{\overline{\mathbf{x}}\cdot\pi}$$

also looks as though it may have such a problem. There are potentially three singularities that may render this expression meaningless:

$$\overline{\mathbf{x}} \perp \pi$$

$$\pi = 0$$

and

$$\frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\Big|_{\overline{\mathbf{x}}=0}$$

This can be sorted out by going back a step from equation 2.10 to equation 2.9

$$H_{(0)} = \frac{m^2c^2\left(1 - \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\right)}{-L_{(0)}}$$

On substituting using equations 2.4 and 1,

$$= m^2c^2\left(1 - \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{\frac{d\mathbf{x}}{dx^0}}{1 + \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{d\mathbf{x}}{dx^0}}\right)\frac{1 + \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{d\mathbf{x}}{dx^0}}{mc\sqrt{1 - \frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}}}$$

$$= mc\frac{1}{\sqrt{1 - \frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}}} \tag{5.6}$$

In this form, it is quite clear that there are no difficult points at all. $H_{(0)}$ bears an uncanny resemblance to the reciprocal of the usual Lagrangian. In fact, rather unexcitingly,

$$H_{(0)} = cp^0 = cp_0 \tag{5.7}$$

$p_0$ being the zero component of the canonical energy–momentum 4–vector.

§§5.4 *The case* $\overline{x}^0 = z^1$

The formula for a trajectory is now

$$\mathbf{x} = \mathbf{x}(\overline{x}^0)$$

or

$$\begin{aligned} x^j &= \overline{x}^j(\overline{x}^0) + z^j \qquad (j = 2,3) \\ x^1 &= \overline{x}^1(\overline{x}^0) + \overline{x}^0 \end{aligned} \tag{5.8}$$

So

$$\frac{dx^j}{da} = \frac{d\overline{x}^j}{d\overline{x}^0}\frac{d\left(x^1 \mp \sqrt{(x^0 - z^0)^2 - (x^2 - z^2)^2 - (x^3 - z^3)^2}\right)}{da} \qquad (j = 2,3)$$

$$\frac{dx^1}{da} = \frac{d\overline{x}^1}{d\overline{x}^0}\frac{d\overline{x}^0}{da} + \frac{d\overline{x}^0}{da}$$

Now

$$\frac{d\overline{x}^0}{dx^0} = \frac{dx^1}{dx^0} \mp \frac{(x^0 - z^0) - (x^2 - z^2)\frac{dx^2}{dx^0} - (x^3 - z^3)\frac{dx^3}{dx^0}}{\sqrt{(x^0 - z^0)^2 - (x^2 - z^2)^2 - (x^3 - z^3)^2}}$$

i.e.,

$$\frac{d\overline{x}^0}{dx^0} = \frac{dx^1}{dx^0} - \frac{-|\overline{\mathbf{x}}| - \overline{x}^2\frac{dx^2}{dx^0} - \overline{x}^3\frac{dx^3}{dx^0}}{\overline{x}^1}$$

So that

$$\frac{d\overline{x}^j}{d\overline{x}^0} = \frac{\frac{dx^j}{dx^0}}{\frac{dx^1}{dx^0} + \left(|\overline{\mathbf{x}}| + \overline{x}^2\frac{dx^2}{dx^0} + \overline{x}^3\frac{dx^3}{dx^0}\right)/\overline{x}^1}$$

$$\frac{d\overline{x}^1}{d\overline{x}^0} = \frac{-\left(|\overline{\mathbf{x}}| + \overline{x}^2\frac{dx^2}{dx^0} + \overline{x}^3\frac{dx^3}{dx^0}\right)/\overline{x}^1}{\frac{dx^1}{dx^0} + \left(|\overline{\mathbf{x}}| + \overline{x}^2\frac{dx^2}{dx^0} + \overline{x}^3\frac{dx^3}{dx^0}\right)/\overline{x}^1}$$

This allows me to re-write the Lagrangian in terms of more familiar entities by substituting the last equations into 2.13:

$$\left(\frac{-L_{(1)}}{mc}\right)^2 = -\left(1+\frac{d\overline{x}^1}{d\overline{x}^0}\right)^2 - \left(\frac{d\overline{x}^2}{d\overline{x}^0}\right)^2 - \left(\frac{d\overline{x}^3}{d\overline{x}^0}\right)^2 + \left(\frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\cdot\frac{d\overline{\mathbf{x}}}{d\overline{x}^0}\right)^2$$

$\Rightarrow$

$$\left(\frac{dx^1}{dx^0}+\frac{|\overline{\mathbf{x}}|+\overline{x}^2\frac{dx^2}{dx^0}+\overline{x}^3\frac{dx^3}{dx^0}}{\overline{x}^1}\right)^2\left(\frac{-L_{(1)}}{mc}\right)^2$$

$$= -\left(\frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}\right)+\left(\frac{\overline{x}^1}{|\overline{\mathbf{x}}|}\left(-\frac{|\overline{\mathbf{x}}|+\overline{x}^2\frac{dx^2}{dx^0}+\overline{x}^3\frac{dx^3}{dx^0}}{\overline{x}^1}\right)+\frac{\overline{x}^2}{|\overline{\mathbf{x}}|}\frac{dx^2}{dx^0}+\frac{\overline{x}^3}{|\overline{\mathbf{x}}|}\frac{dx^3}{dx^0}\right)^2$$

$$= 1-\frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}$$

$\Rightarrow$

$$L_{(1)} = \frac{-mc\sqrt{1-\frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}}}{\frac{dx^1}{dx^0}+\left(|\overline{\mathbf{x}}|+\overline{x}^2\frac{dx^2}{dx^0}+\overline{x}^3\frac{dx^3}{dx^0}\right)/\overline{x}^1}$$

The Hamiltonian, $H_{(1)}$ of equation 2.14, can therefore be re-written:

$$H_{(1)} = \frac{-mc\frac{dx^1}{dx^0}}{\frac{dx^1}{dx^0}+\left(|\overline{\mathbf{x}}|+\overline{x}^2\frac{dx^2}{dx^0}+\overline{x}^3\frac{dx^3}{dx^0}\right)/\overline{x}^1}\;\frac{\frac{dx^1}{dx^0}+\left(|\overline{\mathbf{x}}|+\overline{x}^2\frac{dx^2}{dx^0}+\overline{x}^3\frac{dx^3}{dx^0}\right)/\overline{x}^1}{\sqrt{1-\frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}}}$$

$$= \frac{-mc\frac{dx^1}{dx^0}}{\sqrt{1-\frac{d\mathbf{x}}{dx^0}\cdot\frac{d\mathbf{x}}{dx^0}}} = -cp^1 = cp_1 \qquad (5.9)$$

It should now be blindingly obvious that Dirac was literally correct in his verbal argument: the four Hamiltonians are the four components of the energy–momentum 4–vector.

## §6 Quantum Mechanics on a Backwards–in–Time Light-cone

Certain features are now clear of what must be involved in a quantum mechanics in terms of backward light-cones.

(i) The infinitesimal volume element on the backwards–in–time light-cone is

$$\frac{d^3\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\quad,$$

whence the Hilbert space of states will be $L^2(\Re^3, d^3\overline{\mathbf{x}}/|\overline{\mathbf{x}}|)$;

(ii) the four Hamiltonians that generate evolution from one backward light-cone to another are the four components of the momentum 4–vector, $(p)$;

(iii) any 'configuration space' representation is liable to be based on a transformation from a unitary irreducible representation of the Poincaré group, or at least the orthochronous subgroup.

The most straightforward approach is to posit that configuration space wave functions are defined by

$$\psi(x) = \frac{1}{(2\pi\hbar)^{3/2}} \int_{\Re^3} \frac{d^3\mathbf{p}}{\sqrt[4]{\mathbf{p}\cdot\mathbf{p} + m^2c^2}} \exp\left( -\frac{ix^0\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}}{\hbar} + \frac{i\mathbf{x}\cdot\mathbf{p}}{\hbar} \right) \tilde{\psi}(\mathbf{p}) \qquad (6.1)$$

where, similarly to §2 above,

$$\begin{aligned} x^0 &= z^0 - |\overline{\mathbf{x}}| \\ \mathbf{x} &= \mathbf{z} + \overline{\mathbf{x}} \end{aligned} \qquad (6.2)$$

The 'instantaneous' state of the system is defined by the function $\psi(\overline{\mathbf{x}})$ for a chosen value of the 4–vector $(z)$. To call this instantaneous is, it will be noticed, to considerably broaden the meaning of this word; it now must imply something like, 'a hypersurface in space-time such that if a signal were launched or some action initiated at any point then this would be detectable at $(z)$.' I suspect that the abuse of the word instantaneous is preferable to such phrases.

There is quite a lot, here, of what is usually associated with quantum mechanics. There are also almost as many pitfalls as are commonly found in Poincaré–invariant theories of this kind.

To cope with the singularity in the measure on backward light-cones it is necessary to insist that configuration space wave functions are of the type

$$\psi(x) = \sqrt{|\overline{\mathbf{x}}|}\phi(x) \qquad (6.3)$$

where $\phi \in L^2(\Re^3, d^3\overline{\mathbf{x}})$. This can be incorporated into equation 1 by the inclusion of an extra factor $\sqrt{|\overline{\mathbf{x}}|}$ on the right-hand side. Since, from equation 2, if $(z)$ is a coordinate 4–vector it follows that $(\overline{x}) = (-|\overline{\mathbf{x}}|, \overline{\mathbf{x}})$ has the same transformation properties as a difference of coordinates, and so the factor $\sqrt{|\overline{\mathbf{x}}|} = \sqrt{-\overline{x}^0}$ is not Poincaré–invariant. The solution is that equation 3 is a separate axiom if, in fact, this constraint is compatible with the Poincaré invariance (up to a factor) of $\psi$.

The inverse to formula 1 does not take place on any single backwards–in–time light-cone; it is

$$\frac{\exp\left(-ix^0\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}/\hbar\right)\tilde{\psi}(\mathbf{p})}{\sqrt[4]{\mathbf{p}\cdot\mathbf{p} + m^2c^2}} = \frac{1}{(2\pi\hbar)^{3/2}} \int_{\Re^3} e^{-i\mathbf{x}\cdot\mathbf{p}/\hbar}\psi(x, x^0)\, d^3\mathbf{x} \qquad (6.4)$$

$(x^0$ fixed).

Thus, not only does the configuration space representation look formally similar to the conventional wave function on a constant–$x^0$ hyperplane, the simplest method of producing a state starts from an initial value on a constant–$x^0$ hyperplane. In practical terms this means that the normalisation of states on backward light-cones is not necessarily a consequence of normalisation in the momentum representation. For, consider the equation:

$$\int |\psi(\overline{x})|^2 \frac{d^3\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|} = \int \frac{d^3\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|} \int \frac{d^3\mathbf{p}}{\sqrt{p_0}} \int \frac{d^3\mathbf{p}'}{\sqrt{p'_0}} \tilde{\psi}(\mathbf{p})\tilde{\psi}^*(\mathbf{p}') e^{-ix(p-p')/\hbar} \quad .$$

There seems to be no way to formulate a conservation of probability law for this theory; or even a continuity equation.

The four Hamiltonians can readily be written as differential operators on configuration space:

$$\hat{H}_\nu = i\hbar \frac{\partial}{\partial x^\nu}$$

However, expressing these in terms of the light-cone coordinates is far from easy. This is a direct result of the non-equivalence of the configuration space and momentum space inner-products.

$$-i\hbar \frac{\partial}{\partial \overline{\mathbf{x}}} + i\hbar \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|^2} = -i\hbar\sqrt{|\overline{\mathbf{x}}|} \frac{\partial}{\partial \overline{\mathbf{x}}} \frac{1}{\sqrt{|\overline{\mathbf{x}}|}}$$

is a self-adjoint operator; it is not covariant however (cf. the Newton–Wigner position operator reviewed in Chapter 4).

Computing from formula 1

$$-i\hbar\frac{\partial}{\partial\bar{x}^k} = \frac{\overline{\mathbf{x}}}{|\overline{\mathbf{x}}|}\hat{H}_0 + \hat{H}_k$$

Thus, while there is now a clearly defined position observable (multiplication by $\overline{\mathbf{x}}$), there seems to be no momentum observable.

A *No Go* theorem in the style of Hegerfeldt is the inevitable concomitant of the use of the group representation that gives formula 1. In the presence of such a result there hardly seems much point in trying to hammer out the foregoing catalogue of defects.

## §7 Conclusion

I have produced a revised version of Derrick's 'light-cone mechanics' without noticeably improving on Derrick's wobbly interpretation. By dint of ignoring this and proceeding in a manner more akin to Dirac's actual remarks, a credible version of classical mechanics has been produced.

Attempting to understand the nature of the new dynamical variables, by expressing them in terms of the more usual quantities, is most revealing. Not only is there nothing very new in what is derived – the four Hamiltonians turn out to be the momentum 4–vector – but a variant on the standard quantisation procedure is applicable. The result of this, foreshadowed by the attempted quantisation of the light-cone variables in §4, is that all the problems inherent in the usual treatments of the Klein–Gordon equation remain unaltered (cf. the discussion in chapter 4, §3 and §4).

There are, if anything, more problems with this version of quantum mechanics: there is no problem with the existence of the position observable but now the momentum observable is only partially defined; there is no law of conservation of probability. On reflection, I am left with a wistful affection for the simple, familiar, pitfalls of the conventional approach.

## §8 Bibliography

(References are preceded by the page number on which they first occur.)

2: [1] G. H. Derrick, *Classical mechanics with respect to an observer's past light cone*, J. Math. Phys. 28 (1987) 64.

G. H. Derrick, *Quantum mechanics of a charged scalar boson with respect to an observer's past light cone*, J. Math. Phys. 28 (1987) 1327.

G. H. Derrick, *Complete orthonormal sets on the past light cone. I. Functions belonging to mass zero and helicity s*, J. Math. Phys. 29 (1988) 636–641.

2: [2] P. A. M. Dirac, *Forms of relativistic dynamics*, Rev. Mod. Phys. **21** (1949) 392–9.

2: [3] K. R. Popper, *The Open Universe; an argument for indeterminism*, Hutchinson, 1982; pp 68–77. See also the reference to this same work in chapter 8.

9: [4] I. Stewart, D. Tall, *Complex Analysis*. CUP 1983, p 187.

# Chapter 8

## Feynman's Propagator Approach

*a critique and revision*

*'So, naturalists observe, a flea*
*Hath smaller fleas that on him prey;*
*And these have smaller fleas to bite 'em,*
*And so proceed* ad infinitum.
*Thus every poet, in his kind,*
*Is bit by him that comes behind.'*

Jonathan Swift **On Poetry.**

## §1 Introduction

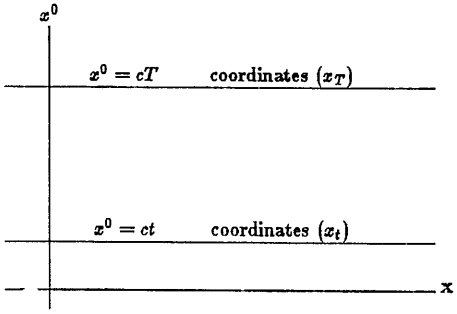
In considering the problem of localisation in quantum mechanics it is obviously necessary to review all the notable efforts made in the past. With the exception of Feynman's work, it is my hope that the previous chapters do justice to this task. Feynman's quantum electrodynamics (QED)[1] is far and away the most important of these 'notable efforts', despite the fact that this problem is never directly addressed. There are a couple of reasons for this omission: from the publication dates and peripheral records[2] it is apparent that Feynman's main QED papers (1948–51) largely pre-date the localisation problem – if this is set, *pace* historians, in 1949 with the paper by Newton and Wigner. Also, since his main objective was always to calculate experimentally verifiable numbers, it seems unlikely that this more metaphysical problem would have caught his eye. His is a theory that reaches well beyond the confines of the localisation problem – something rarely the case with the other authors reviewed. It is also a theory with a tremendous body of experimental support behind it.

Quantum electrodynamics must be radically simplified before such burning issues as 'causality' and locality can be examined, though this is accomplished with a single, additional constraint on the 4-potential. This simplification discards (retrievably) virtually all the interesting and useful content of the theory. It does, however, remove the tricky issue of renormalisation.

In subsequent sections of this chapter the reduced form of QED, now a single-particle quantum mechanics, will be summarised, criticised and an improvement offered. The criticism stems from the mathematical observations of authors such as Lars Gårding,[3] Daniel Zwanziger,[4] Lars Hörmander,[5], Gerhard Hegerfeldt,[6] and others: there are difficulties inherent in formulating Poincaré-invariant theories that are 'causal' (as defined in Chapter 5). The Einstein–Podolsky–Rosen thought-experiment and the Double Slit experiment are also discussed, as illustrations of the new theory and its interpretation.

QED theory is based on constructs called 'propagators' or 'Green's functions'. The primary advantage of the propagator approach is that amplitudes are computed between pairs of space-time events. By beginning with these more elementary connexions, rather than some hypersurface, there is a chance that the objections of previous chapters may be avoided. It is true that, in the usual elaboration of a theory based on propagators, constant-time hyperplanes are introduced as the location of initial values. This is not, fortunately, the only way of solving a wave equation, which can also be tackled as a boundary-value problem.

The variation on Feynman's successful formalism offered is the analogue to a formulation of the quantum mechanics of the photon. A number of features recommend this free-electron propagator, not least its close similarity to Feynman's.




## §2 A Derivation And Review Of Feynman's Approach

To obtain a one-particle quantum mechanics from Feynman's multiple-particle QED requires only a condition on the 4–potential:

$$A_\mu = 0 \tag{2.1}$$

the reason for this is that, as Feynman put it, all the creation and annihilation of the multitude of particles occurs by scattering at the various points of the applied electro-magnetic field. Thus a universe containing a single electron and obeying the foregoing constraint will only ever contain that electron. What remains of QED is, therefore, the free evolution of a single electron. This evolution is realised through the good services of a propagator.

Consider the region of space-time between two constant-time hyperplanes ( $ct < x^0 < cT$ ).

The kernel of an integral operator, denoted by $K(x,y)$ , is called a *propagator* if, given that the amplitude of the particle is $\phi(y)$ at the event, $(y)$ , then this contributes

$$K(x,y)\phi(y)$$

to the amplitude of the particle at the space-time event, $(x)$ .

Feynman proposed that if the system has the electron state $\phi(x_t)$ (a positive energy spinor) on the $x^0 = ct$ hyperplane, and the positron state $\phi(x_T)$ (a negative energy spinor) on the $x^0 = cT$ hyperplane, then at any event, $(x)$ , such that $x^0 \in (ct, cT)$ , the system will have the amplitude $\phi(x)$ given by

$$\phi(x) = \int K(x,x_t)\phi(x_t)\,d^3x_t - \int K(x,x_T)\phi(x_T)\,d^3x_T \quad . \tag{2.2}$$

Thus the amplitude at $(x)$ is the difference between the compounded amplitude that there was an electron in state $\phi(x_t)$ in the past and that there will be a positron in state $\phi(x_T)$ in the future. To accomplish this, conditions are placed on the propagator so that if $(x)$ is an event in the future of $(y)$ then $K(x,y)$ will contribute an amplitude having only positive energy components, and (swapping $x$ and $y$ ) $K(y,x)$ produces only negative energy components.

To derive Feynman's free-electron propagator, start from equation 2 and set $\phi(x_T) = 0$ :

$$\theta(x^0 - ct)\phi(x) = \int K(x,x_t)\phi(x_t)\,d^3x_t \quad . \tag{2.3}$$

The Heaviside unit-step function, $\theta$ , is necessary because $\phi$ represents an electron – to obtain $\phi(x)$ for $x^0 < ct$ a new, prior, initial-value must be given.

Now, if $\phi$ is the state of an electron it must satisfy Dirac's equation:

$$\left(i\hbar\gamma^\mu \frac{\partial}{\partial x^\mu} - mc\right)\phi = 0 \quad . \tag{2.4}$$

Applying the differential operator above to both sides of equation 3 then gives

$$\phi(x)i\hbar\gamma^0\frac{\partial\theta(x^0-ct)}{\partial x^0}=\int\left(i\hbar\gamma^\mu\frac{\partial}{\partial x^\mu}-mc\right)K(x,x_t)\phi(x_t)\,d^3x_t \quad . \tag{2.5}$$

However, $\theta$ is a generalised function,[7] and, as such, is only properly defined as one half of an integrand. Given some suitable function, $f$ , then

$$\int_{-\infty}^{+\infty}\theta(x)f(x)\,dx=\int_{0}^{+\infty}f(x)\,dx=F(\infty)-F(0) \tag{2.6}$$

may be said to be well-defined. Sufficient conditions for what follows are

$$\int_{-\infty}^{x}f(y)\,dy=F(x)<\infty$$

and

$$F\in L^1(\Re) \quad .$$

Integrating the left-hand side of equation 6 by parts gives, since $F(\infty)=0$,

$$-F(0)=\left[\theta(x)F(x)\right]_{x=-\infty}^{\infty}-\int_{-\infty}^{+\infty}\frac{\partial\theta}{\partial x}F(x)\,dx$$

i.e.,

$$F(0)=\int_{-\infty}^{+\infty}\frac{\partial\theta}{\partial x}F(x)\,dx$$

so that, as a generalised function,

$$\frac{\partial\theta}{\partial x}=\delta(x) \quad . \tag{2.7}$$

Thus equation 5 may be written

$$\int_{\Re^3}\left(i\hbar\gamma^\mu\frac{\partial}{\partial x^\mu}-mc\right)K(x,x_t)\phi(x_t)\,d^3x_t=i\hbar\gamma^0\delta(x^0-ct)\phi(x)$$

which can be written as an equation of integrals:

$$=i\hbar\gamma^0\int_{\Re^3}\delta_4(x-x_t)\phi(x_t)\,d^3x_t$$

where $\delta_4(x)=\delta(x^0)\delta(x^1)\delta(x^2)\delta(x^3)$ .

The conclusion that can be drawn from this is, then, in terms of generalised functions at least,

$$\left(i\hbar\gamma^\mu\frac{\partial}{\partial x^\mu}-mc\right)K(x,x_t)=i\hbar\gamma^0\delta_4(x-x_t) \quad . \tag{2.8}$$

This equation, together with a boundary condition, determines the free-electron propagator. The boundary condition imposed by Feynman is that only positive energy components should propagate forwards in time: this has the immediate consequence that only negative energy components propagate backwards in time. To solve this equation it is necessary to pass to the conjugate, 4-momentum space, where

$$(\gamma^\mu p_\mu-mc)\tilde{K}(p,x_t)=\frac{i\hbar\gamma^0}{(2\pi\hbar)^2}e^{ip_\nu x_t^\nu/\hbar}$$

Multiplying through by

$$\frac{\gamma^\mu p_\mu+mc}{p^\nu p_\nu-m^2c^2}$$

gives, since the $\gamma^\mu$ anti-commute,

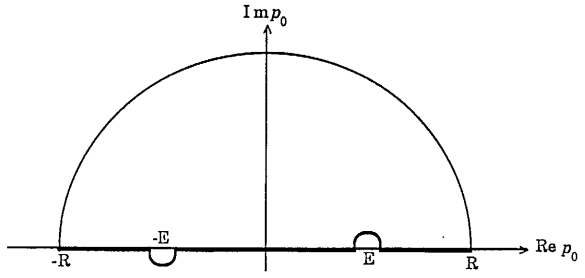
$$\tilde{K}(p, x_t) = \frac{\gamma^\mu p_\mu + mc}{p^\nu p_\nu - m^2c^2} \frac{i\hbar\gamma^0}{(2\pi\hbar)^2} e^{ip_\nu x_t^\nu/\hbar} \quad . \tag{2.9}$$

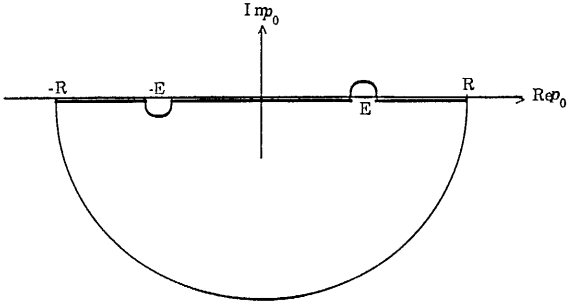
To perform the inverse Fourier transform, it is easiest to find first the inverse transform between $p_0$ and $x^0$:

$$\tilde{K}(\mathbf{p}, x^0, x_t) = e^{-i\mathbf{p}\cdot\mathbf{x}_t/\hbar} \int_{-\infty}^{\infty} \frac{\gamma^\mu p_\mu + mc}{p_0^2 - (\mathbf{p}\cdot\mathbf{p} + m^2c^2)} e^{ip_0(ct - x^0)/\hbar}\, dp_0 \, \frac{i\hbar\gamma^0}{(2\pi\hbar)^{5/2}} \quad . \tag{2.10}$$

There are two simple poles in this integration— at $p_0 = \pm\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}$ —and it is through the choice of the complex contour used to perform the integral that Feynman's boundary condition is met. Two contours are actually used, depending on the sign of $ct - x^0$. (For simplicity, let $E = \sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}$ in what follows.)

$ct - x^0 > 0$




$ct - x^0 < 0$

The semicircles around $p_0 = \pm E$ are taken to be of arbitrarily small radius, $r$ ; and once the residue theorem has been applied the limit $r \to 0$ can be assumed.

**Re-writing the integral as**

$$\int_{-\infty}^{\infty} \frac{\gamma^\mu p_\mu + mc}{2E} e^{ip_0(ct-x^0)/\hbar} \left( \frac{1}{p_0 - E} - \frac{1}{p_0 + E} \right) dp_0 \quad , \tag{2.11}$$

**it is then a simple matter to deduce, via Cauchy's residue theorem, that**

$$\begin{aligned} &\text{if } ct - x^0 > 0 \text{ the integral is } 2\pi i \frac{-E\gamma^0 - \boldsymbol{\gamma}\cdot\mathbf{p} + mc}{-2E} \exp\left( \frac{-iE(ct - x^0)}{\hbar} \right) \\ &\text{and} \\ &\text{if } ct - x^0 < 0 \text{ the integral is } 2\pi i \frac{E\gamma^0 - \boldsymbol{\gamma}\cdot\mathbf{p} + mc}{2E} \exp\left( \frac{iE(ct - x^0)}{\hbar} \right) \end{aligned} \tag{2.12}$$

These contour integrals give the required integral along the real axis by taking the limit $R \to \infty$, since the residue theorem gives the same result for all values of $R$ but the integral around the semicircle of radius $R$ gives an ever-decreasing contribution:

Set $p_0 = Re^{i\theta}$, and change the variable of integration on the semicircle to $\theta$ :

$$dp_0 = iRe^{i\theta}\, d\theta$$

whence the integral around the semicircle becomes

$$\int_{\theta_1}^{\theta_2} \exp\left( R\frac{(ct - x^0)}{\hbar}(i\cos\theta - \sin\theta) \right) \frac{\gamma^0 Re^{i\theta} - \boldsymbol{\gamma}\cdot\mathbf{p} + mc}{R^2 e^{2i\theta} - (\mathbf{p}\cdot\mathbf{p} + m^2c^2)} iRe^{i\theta}\, d\theta$$

This integrand behaves, in modulus, as

$$\exp\left( -\frac{R(ct - x^0)\sin\theta}{\hbar} \right) \tag{2.13}$$

Now, since $R > 0$, this integral will vanish in the limit $R \to \infty$ if

$$-(ct - x^0)\sin\theta < 0 \quad . \tag{2.14}$$

Thus, if $ct - x^0 > 0$ take $\theta \in [0, \pi]$ ; if $ct - x^0 < 0$ take $\theta \in [-\pi, 0]$ . Gathering the result in formula 12 into equation 10, and taking the inverse Fourier transforms with respect to $\mathbf{p}$ gives

$$\begin{aligned} K(x, x_t) = \frac{(\gamma^\mu p_\mu + mc)(-\gamma^0)}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} \frac{d^3\mathbf{p}}{2\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}} e^{i\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}_t)/\hbar} \times \\ \left[ \theta(x^0 - ct) \exp\left( i\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}\frac{(ct - x^0)}{\hbar} \right) - \right. \\ \left. \theta(ct - x^0) \exp\left( -i\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}\frac{(ct - x^0)}{\hbar} \right) \right] \quad . \end{aligned} \tag{2.15}$$

This generalised function behaves exactly as advertised. To progress further in specifying its action would now require the introduction of a specific example. The presentation of Feynman's theory is, therefore, accomplished, for my purposes.

## §3 A Critique

**Quantum electrodynamics, despite its successes, has a number of problems: renormalisation and virtual particles, to name the obvious two. It is my contention that even when I obviate these problems – by setting the 4–potential to zero – important difficulties remain in producing a workable quantum mechanics (a first quantised theory).**

The simplest of these difficulties is the noninvariance of the propagator: take two events, $(x)$ and $(y)$ , that are space-like separated. If $y^0 < x^0$ then the electron amplitude, $\phi$ , induced at $(x)$ by the amplitude at $(y)$ is

$$K(x,y)\phi(y)$$

But, because these events are space-like separated, there is a Lorentz boost to a new frame of reference that reverses their time order – so that $\overline{x}^0 > \overline{x}^0$. In this coordinate frame

$$K(\overline{x},\overline{y})\phi(\overline{y})$$

can only relate the positron amplitude at $(\overline{y})$ to the positron amplitude at $(\overline{x})$ : by the formulation outlined in the last section there is no connection of electron amplitudes. This is entirely the fault of the functions $\theta(x^0 - ct)$ and $\theta(ct - x^0)$ in the formula 2.15 that was found for $K(x, x_t)$ . This, in turn, is the consequence of the contours chosen by Feynman – the 'Feynman rules'.

A second difficulty arises from the application of the Paley–Wiener theorem (cf. Chapter 5). Simply put, an electron (or positron) state that is initially localised – has a compact spatial support – will at no future (respectively, past) time have a finite support. This obviously implies that the amplitude spreads in a manner heedless of the speed of light, which might be thought inconsiderate, but there you are.

The application of the Paley–Wiener theorem runs as follows: take four functions, $\psi_i \in L^2(\Re^3)$ such that

$$(\forall i)(\exists r : 0 < r < \infty) \qquad \text{supp}\,\psi_i \subset B(\mathbf{0};r)$$

(compact support). Form these into a Dirac spinor, $\Psi$ . A Fourier transform, $\mathcal{F}_4$ is defined on this spinor in the obvious manner: as a component-by-component transform

$$\mathcal{F}_4 = I_4\mathcal{F}_1 = \begin{pmatrix} \mathcal{F}_1 & 0 & 0 & 0 \\ 0 & \mathcal{F}_1 & 0 & 0 \\ 0 & 0 & \mathcal{F}_1 & 0 \\ 0 & 0 & 0 & \mathcal{F}_1 \end{pmatrix} \qquad \mathcal{F}_1 = \frac{1}{(2\pi\hbar)^{3/2}}\int_{\Re^3} d^3\mathbf{y}\, e^{-i\mathbf{p}\cdot\mathbf{y}/\hbar} \times \quad .$$

If $\Psi$ is taken to be the initial state of an electron then its evolution to all future times may be written as

$$\Psi(x) = (\gamma^\mu p_\mu + mc)(-\gamma^0)\mathcal{F}^{-1}\left[\frac{e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})/\hbar}}{2\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}}\exp\left(\frac{i(y^0-x^0)}{\hbar}\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}\right)\mathcal{F}\left[\Psi(y^0,\mathbf{y})\right]\right]$$

Now this is a consistent equation, so that setting $x^0 = y^0$ recovers the initial condition on the left-hand side. What this means is, re-arranging this formulation of the propagator, and applying the Paley–Wiener theorem, that

$$f(\mathbf{p}) = \frac{(\gamma^\mu p_\mu + mc)(-\gamma^0)}{2\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}}\mathcal{F}\left[\Psi(y^0,\mathbf{y})\right]$$

is an exponential entire function. However, because of the square root, the function

$$\exp\left(\frac{i(y^0-x^0)}{\hbar}\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}\right)f(\mathbf{p})$$

is not analytic, and therefore not exponential entire. By the contra-positive of the Paley–Wiener theorem, it follows that, for $x^0 \neq y^0$ ,

$$\mathcal{F}^{-1}\left[\exp\left(\frac{i(y^0-x^0)}{\hbar}\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2}\right)f(\mathbf{p})\right]$$

**is not a function of compact spatial support. The locality problem is thereby manifest for Feynman's theory of the electron.**

**Now it is possible to introduce a positron state in such a way that the resulting amplitude is of compact spatial support between the two initial values. This does, however, seem to be a precariously contrived way to rescue locality; indeed, somehow, the positron must escape detection every time.**

**Of course, there is no reason to think that my judgement of these difficulties is final and fatal. It could be argued, just as many of the authors reviewed in Chapter 4 have argued, that these are excusable faults. That the first difficulty is an artefact of the production of a probability density that is the 0–component of a probability 4–current. The second difficulty can be called into question by the observation that Feynman's theory is in such excellent agreement with experiment. It is my purpose to thoroughly de-bunk such disingenious cavils in the remainder of this chapter.**

## §4 The Quantum Mechanics Of 'The Photon'

There are two reasons for examining at this point what might be called the quantum mechanics of the photon, first: the mathematics is simpler and readily solved; and second: there is no dispute about the way photons move – at all times at the universal constant speed of light. In the main, this analysis marks a useful step in the reformulation of Feynman's theory. To begin with the equivalent of the development of §2 will be performed, but a crucial change will be made on the basis of arguments analogous to those of the previous section.

The wave equation for a mass-less scalar particle is the three-dimensional Wave Equation:

$$\hbar^2 \left( \frac{\partial^2}{(\partial x^0)^2} - \nabla^2 \right) \psi = 0 \quad . \tag{4.1}$$

As before, the propagator is sought, which, in this case, satisfies

$$\hbar^2 \left( \frac{\partial^2}{(\partial x^0)^2} - \nabla_x^2 \right) G(x,y) = \delta_4(x-y) \quad . \tag{4.2}$$

The four-dimensional Fourier transform of this, with respect to $x$ , is therefore the solution of

$$-(p_\nu p^\nu)\tilde{G}(p,y) = e^{ip_\nu y^\nu/\hbar} \quad . \tag{4.3}$$

Thus

$$G(x,y) = \frac{1}{(2\pi\hbar)^4} \int_{\mathbb{R}^4} d^4p \frac{e^{ip_\nu \xi^\nu/\hbar}}{|\mathfrak{p}|^2 - p_0^2} \qquad \xi^\nu = y^\nu - x^\nu$$

Changing to spherical polar coordinates:

$$\vec{p} = \sqrt{\mathfrak{p} \cdot \mathfrak{p}}$$

$$\theta = \cos^{-1}\left( \frac{\mathfrak{p} \cdot \vec{\xi}}{|\mathfrak{p}||\vec{\xi}|} \right)$$

$$\phi = \tan^{-1}\left( \frac{p_2'}{p_1'} \right)$$

where $(\mathfrak{p}')$ are the Cartesian coordinates of $\mathfrak{p}$ in which $p_3'$ is parallel to $\vec{\xi}$ . Thence,

$$\mathfrak{p} \cdot \vec{\xi} = \vec{p}\xi \cos\theta \qquad (\xi = |\vec{\xi}|)$$

and

$$G(x,y) = \frac{1}{(2\pi\hbar)^4} \int_{-\infty}^{\infty} dp_0 \int_0^{\infty} (\vec{p})^2 \, d\vec{p} \int_0^{\pi} \sin\theta \, d\theta \int_0^{2\pi} d\phi \frac{e^{-i(\vec{p}\xi\cos\theta - p_0\xi^0)/\hbar}}{(\vec{p})^2 - p_0^2}$$

The $\theta$- and $\phi$–integrals are readily computed:

$$\frac{\partial e^{-i\vec{p}\vec{\xi}\cos\theta/\hbar}}{\partial\theta} = \frac{i\vec{p}\vec{\xi}\sin\theta}{\hbar} e^{-i\vec{p}\vec{\xi}\cos\theta/\hbar}$$

whence

$$G(x,y) = \frac{2\pi}{(2\pi\hbar)^4}\int_{-\infty}^{\infty} dp_0 \int_0^{\infty} (\vec{p})^2\, d\vec{p}\, \frac{e^{ip_0\xi^0/\hbar}}{(\vec{p})^2 - p_0^2} \frac{e^{i\vec{p}\vec{\xi}/\hbar} - e^{-i\vec{p}\vec{\xi}/\hbar}}{i\vec{p}\vec{\xi}/\hbar}$$

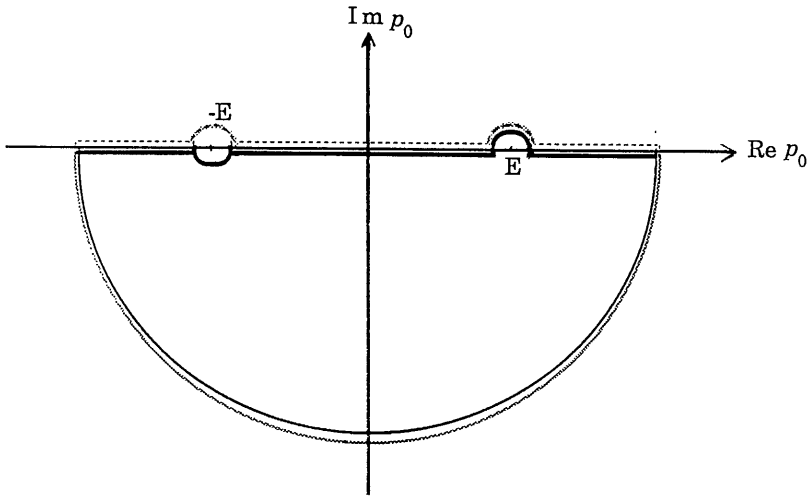
i.e.,

$$G(x,y) = \frac{4\pi}{(2\pi\hbar)^4\xi}\int_0^{\infty} d\vec{p} \int_{-\infty}^{\infty} dp_0 \frac{\vec{p}e^{ip_0\xi^0/\hbar}}{(\vec{p})^2 - p_0^2} \sin\left(\frac{\vec{p}\vec{\xi}}{\hbar}\right) \quad . \tag{4.4}$$

As in §2, the $p_0$–integral can only be evaluated as a complex contour integral, the choice of contour being the tricky bit. Of the four possibilities available for each sign of $\xi^0$ only two really merit closer scrutiny – the four options can be most easily described in terms of the poles included (they occur at $p^0 = \pm\vec{p}$ ); in full, the set is

$$\{\text{none}, +\vec{p}, -\vec{p}, \text{both}\}$$

– taking $\xi^0 < 0$ , these are illustrated next.




The sensible alternatives are: $+\vec{p}$ only—Feynman's rule—or both $+\vec{p}$ and $-\vec{p}$ —conventionally termed the retarded solution. Obviously the use of neither pole is uninteresting, and it seems as ridiculous to include only the $-\vec{p}$ pole when $\xi^0 < 0$ ( $(x)$ in the future of $(y)$ ).

Now

$$\frac{e^{ip_0\xi^0/\hbar}}{(\vec{p})^2 - p_0^2} = \frac{1}{2\vec{p}}\left(\frac{e^{ip_0\xi^0/\hbar}}{p_0 + \vec{p}} - \frac{e^{ip_0\xi^0/\hbar}}{p_0 - \vec{p}}\right)$$

whence the residues at the poles are

$$p_0 = \vec{p} \quad \text{giving} \quad -\frac{e^{i\vec{p}\xi^0/\hbar}}{2\vec{p}} \qquad : \qquad p_0 = -\vec{p} \quad \text{giving} \quad \frac{e^{-i\vec{p}\xi^0/\hbar}}{2\vec{p}}$$

Sticking with $\xi^0 < 0$ , this gives

$$G(x,y) = \begin{cases} -\frac{4\pi^2 i}{(2\pi\hbar)^4 \xi} \int_0^\infty d\vec{p}\, e^{i\vec{p}\xi^0/\hbar} \sin(\vec{p}\xi/\hbar) & \text{(Feynman's rule)} \\ \frac{8\pi^2}{(2\pi\hbar)^4 \xi} \int_0^\infty d\vec{p} \sin(\vec{p}\xi/\hbar) \sin(\vec{p}\xi^0/\hbar) & \text{(retarded/advanced solution)} \end{cases}$$

Using $e^{\pm i\theta} = \cos\theta \pm i\sin\theta$ , this may be written as

$$G(x,y) = \begin{cases} \frac{4\pi^2}{(2\pi\hbar)^4 \xi} \int_0^\infty d\vec{p} \sin(\vec{p}\xi/\hbar) \left(\sin(\vec{p}\xi^0/\hbar) - i\cos(\vec{p}\xi^0/\hbar)\right) & \text{(Feynman's rule } \xi^0 < 0\text{)} \\ \frac{4\pi^2}{(2\pi\hbar)^4 \xi} \int_0^\infty d\vec{p} \sin(\vec{p}\xi/\hbar) \left(\sin(\vec{p}\xi^0/\hbar) + i\cos(\vec{p}\xi^0/\hbar)\right) & \text{(Feynman's rule } \xi^0 > 0\text{)} \\ \frac{8\pi^2}{(2\pi\hbar)^4 \xi} \int_0^\infty d\vec{p} \sin(\vec{p}\xi/\hbar) \sin(\vec{p}\xi^0/\hbar) & \text{(retarded/advanced solution)} \end{cases}$$

There are, consequently, two integrals of interest

$$\int_0^\infty d\vec{p} \sin\left(\frac{\vec{p}\xi}{\hbar}\right) \sin\left(\frac{\vec{p}\xi^0}{\hbar}\right) \quad \text{and} \quad \int_0^\infty d\vec{p} \sin\left(\frac{\vec{p}\xi}{\hbar}\right) \cos\left(\frac{\vec{p}\xi^0}{\hbar}\right)$$

This can be simplified using the trigonometric identities

$$\sin\left(\frac{\vec{p}\xi}{\hbar}\right) \sin\left(\frac{\vec{p}\xi^0}{\hbar}\right) = \frac{1}{2}\left(\cos\frac{\vec{p}}{\hbar}(\xi^0 - \xi) - \cos\frac{\vec{p}}{\hbar}(\xi^0 + \xi)\right) \quad :$$

$$\sin\left(\frac{\vec{p}\xi}{\hbar}\right) \cos\left(\frac{\vec{p}\xi^0}{\hbar}\right) = \frac{1}{2}\left(\sin\frac{\vec{p}}{\hbar}(\xi - \xi^0) + \sin\frac{\vec{p}}{\hbar}(\xi + \xi^0)\right) \quad .$$

Substituting $k = \vec{p}/\hbar$ . and writing $\chi$ for $\xi^0 - \xi$ or $\xi^0 + \xi$ , the integrals of interest become

$$\pm\hbar \int_0^\infty \cos k\chi\, dk \quad \text{and} \quad \hbar \int_0^\infty \sin k\chi\, dk \quad .$$

By all conventional wisdom these integrals do not exist. Once more, generalised functions gallop to the rescue. The generalised function, $\delta$ , is usually represented by the equation

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ipx}\, dp$$

But this is an integral over an even interval, so only the even part of the integrand contributes:

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos px\, dp$$

i.e.,

$$\delta(x) = \frac{1}{\pi} \int_0^\infty \cos px\, dp \quad .$$

This clears up one half of the problem. The sine integral is a little more tricky: the one clear-cut property of this integral is that for $\chi = 0$ it vanishes. Now

$$\sin k\chi = \cos\left(k\chi - \frac{\pi}{2}\right) = -\cos\left(k\chi + \frac{\pi}{2}\right)$$

So, just to be symmetric, try the trigonometric identity

$$\sin k\chi = \frac{1}{2}\left(\cos\left(k\chi - \frac{\pi}{2}\right) - \cos\left(k\chi + \frac{\pi}{2}\right)\right)$$

Thus

$$\int_0^\infty \sin k\chi \, dk = \frac{1}{2}\int_0^\infty \cos\left(k\chi - \frac{\pi}{2}\right) - \cos\left(k\chi + \frac{\pi}{2}\right) dk$$

Making use of the substitutions $K = k - \pi/(2\chi)$ and $K = k + \pi/(2\chi)$

$$\int_0^\infty \sin k\chi \, dk = \frac{1}{2}\left(\int_{-\pi/(2\chi)}^\infty \cos K\chi \, dK - \int_{\pi/(2\chi)}^\infty \cos K\chi \, dK\right)$$

i.e.,

$$= \frac{1}{2}\int_{-\pi/(2\chi)}^{\pi/(2\chi)} \cos K\chi \, dK$$

which is defined unless $\chi = 0$, but the value of the integral at this point is already known, so all that is needed is to evaluate this simple, definite integral:

$$\hbar\int_0^\infty \sin k\chi \, dk = \frac{\hbar}{2}\left[\frac{\sin K\chi}{\chi}\right]_{K=-\pi/(2\chi)}^{\pi/(2\chi)}$$

$$= \frac{\hbar}{2\chi}(1-(-1)) = \frac{\hbar}{\chi} \quad .$$

Recalling that

$$\pm\hbar\int_0^\infty \cos k\chi \, dk = \pm\pi\hbar\delta(\chi) \quad ,$$

I can now gather the results I have amassed to write

$$G(x,y) = \begin{cases} \frac{4\pi^2}{(2\pi\hbar)^4\xi}\left[\frac{1}{2}\left(\pi\hbar\delta(\xi^0-\xi) - \pi\hbar\delta(\xi^0+\xi)\right) - \frac{i\hbar}{2}\left(\frac{1}{\xi+\xi^0}+\frac{1}{\xi-\xi^0}\right)\right] & \\ & \text{(Feynman's rule } \xi^0 < 0\text{)} \\ \frac{4\pi^2}{(2\pi\hbar)^4\xi}\left[\frac{1}{2}\left(\pi\hbar\delta(\xi^0-\xi) - \pi\hbar\delta(\xi^0+\xi)\right) + \frac{i\hbar}{2}\left(\frac{1}{\xi+\xi^0}+\frac{1}{\xi-\xi^0}\right)\right] & \\ & \text{(Feynman's rule } \xi^0 > 0\text{)} \\ \frac{4\pi^2}{(2\pi\hbar)^4\xi}\left(\pi\hbar\delta(\xi^0-\xi) - \pi\hbar\delta(\xi^0+\xi)\right) & \text{(retarded/advanced solution)} \end{cases}$$

Since, if $\phi$ has simple roots $\{a_i\}$,

$$\delta(\phi(x)) = \sum_i \frac{1}{\phi'(a_i)}\delta(x-a_i)$$

then $G(x,y)$ can be written more succinctly as

$$G(x,y) = \begin{cases} \frac{\pi^2}{(2\pi\hbar)^3}\delta\left((\xi^0)^2-\xi^2\right) - \frac{i\pi}{(2\pi\hbar)^3}\frac{1}{\xi^2-(\xi^0)^2} & \text{(Feynman's rule } \xi^0 < 0\text{)} \\ \frac{\pi^2}{(2\pi\hbar)^3}\delta\left((\xi^0)^2-\xi^2\right) + \frac{i\pi}{(2\pi\hbar)^3}\frac{1}{\xi^2-(\xi^0)^2} & \text{(Feynman's rule } \xi^0 > 0\text{)} \\ \frac{2\pi^2}{(2\pi\hbar)^3}\delta\left((\xi^0)^2-\xi^2\right) & \text{(retarded/advanced solution)} \end{cases}$$

It is clear that Feynman's rule does not give 'causal' propagators – the second term is non-zero everywhere except on the light-cone. It is entirely understandable that the advanced/retarded solution for $G(x,y)$ is the one used in practice, well, at least outside QED. Clearly, the use of negative frequencies for the forwards-in-time part of the propagator must be explained, after all it is by excluding this that Feynman's rule comes about.

The next section will take up these points for massive particles.

## §5 A Variation On Propagators

**§§5.1 ... *the story so far***

**In the foregoing sections several important conclusions have been drawn:**

**(i) Feynman's rule produces a non-invariant formalism;**

**(ii) there is also a violation of the Special Principle of Relativity ('causality');**

**(iii) similar problems occur if Feynman's rule is applied to the photon, but this is not the only possible 'rule', and a more credible one has been found;**

**(iv) satisfaction of the Special Principle of Relativity hinges on the use of negative frequencies in the forwards-in-time part of the propagator (and the use of positive frequencies in the backwards-in-time part).**

**The first thing to do is to adopt an invariant formalism. Since the space-time subsets that are** Poincaré-invariant are

(a) the entirety of space-time, $\mathfrak{R}^4$ ;

(b) individual events, $(x)$ ;

it is reasonable, therefore, that constant-time hyperplanes should play little or no part in the theory. This may seem a bit drastic; after all, where is the initial state to be? What of conservation of probability? How are measurements to be represented now? I believe that I have satisfactory answers to these questions in what follows.

§§5.2 *A new free-electron propagator*

In place of an initial state, for the usual initial-value problem, a source function is employed as the inhomogeneous part of an inhomogeneous boundary-value problem with homogeneous boundary conditions (i.e., zero). For the electron this is written

$$\left(i\hbar\gamma^\mu \frac{\partial}{\partial x^\mu} - mc\right)\psi(x) = S_+(x) \tag{5.1}$$

$$\text{supp}\, S_+ = B \qquad (\ B \text{ some subset of space-time})$$

Boundary condition:

$$\psi(x_t, \mathbf{x}) = 0 \qquad \frac{\partial\psi}{\partial x^0}(x_t, \mathbf{x}) = 0$$

The constant-time hyperplane, $x^0 = x_t$ , is taken to be in the past of $B$ —this is only necessary for the solution to be unique and to be the result of a particle 'created' entirely in $B$ .

The Green's function for this problem is given by a remarkably similar equation to Feynman's (equation 2.8)

$$\left(i\hbar\gamma^\mu \frac{\partial}{\partial x^\mu} - mc\right) K(x, x_t) = \delta_4(x - x_t) I_4 \quad .$$

Though now the amplitude is given by

$$\psi(x) = \int_B K(x, y) S_+(y)\, d^4y$$

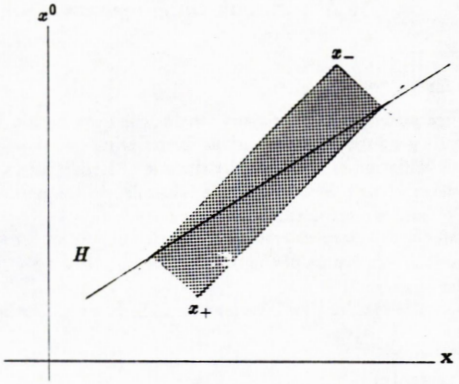
**– an entirely invariant formula.**

Contours analogous to those leading to the advanced/retarded solution for the photon are employed, that is, including both poles within the contour for each sign of $x^0 - y^0$. Then changing to spherical polar coordinates, again as in the last section, gives instead of equation 15 of §2,

$$K(x, x_t) = (\gamma^\mu \hat{p}_\mu + mc)\frac{4\pi}{(2\pi\hbar)^3 \xi} \int_0^\infty \frac{\vec{p}}{\sqrt{(\vec{p})^2 + m^2c^2}} \sin\left(\frac{\vec{p}\xi}{\hbar}\right) \sin\left(\frac{\xi^0}{\hbar}\sqrt{(\vec{p})^2 + m^2c^2}\right) d\vec{p} \quad . \tag{5.2}$$

**The limit $m \to 0$ recovers the case of zero rest mass (the spin-half version of the photon).**

Since this is the kernel of the inhomogeneous boundary-value problem the result conforms to the Special Principle – it is 'causal' – as the characteristic surface of the Dirac equation is the light-cone.[7]




§§5.3 *A justification of 'negative energies'*

A highly desirable result has been obtained – call it 'causality', 'locality', or what-you-will – but was the price too high? The 'negative frequency' fundamental solutions

$$u_\sigma(p)\exp\left(\frac{i}{\hbar}(\mathbf{p}\cdot\mathbf{x}+x^0\sqrt{\mathbf{p}\cdot\mathbf{p}+m^2c^2})\right)$$

are also called the negative energy solutions, and for this reason are usually thought to be unphysical, there being no sensible instance in which a negative energy has been detected. Needless to say, I have an answer to this.

The justification is that there is no reason to suppose that a negative frequency wave has negative energy. The direct connection relies upon the use of an energy operator ('observable') which has negative eigen-values for these negative frequency solutions. As I will presently argue, the basis for associating an operator with an 'observable', and thence a physical quantity, is not necessarily complete, and, in my view, is to be rejected in the case of negative energies (this depends on a model of experimental measurements to be outlined below).

A second reason for insisting on negative frequencies is to make the theory more time-symmetric: just as the state in the past determines that of the future, the reverse must also be the case. This is often grandiloquently labelled 'prediction' and 'retrodiction', terms I will avoid as they imply far more than is sensibly the case.[8] Feynman and Stückelberg used negative frequencies as the means by which positrons determine the state of their prior selves.[9] As an interpretational device this is reasonable: all I want to do is combine it with the positive frequency interpretation into a single expression rather than insist that, somehow, the propagator has an arrow built-in which is compared to the difference of the time-like coordinates (the non-invariant $\theta$-functions).

§§5.4 *A source for the goose is a sink to the gander*

In pursuit of complete time-symmetry, it is obviously necessary to oppose a sink function, $S_-$ , to the source function, $S_+$ . Just as the support of the amplitude spreads through the forward light-cone of the support of $S_+$ , so too the amplitude must converge within the backward light-cone of $S_-$ . Since $S_-$ may be arbitrarily chosen, an algorithm applicable to any $S_-$ must be found for the calculation of amplitudes. Take two events:

$$(x_+) \in \operatorname{supp} S_+$$
$$(x_-) \in \operatorname{supp} S_-$$

then some amplitude will be deployed between this pair. To investigate this it will be assumed first that these are the only elements of their respective supports – as though the 'creation' and 'annihilation' of the particle occur in such small regions that they can be treated as geometric points.

On the hyperplane, $H$ , the amplitude, $\phi(x_H)$ , will be

$$K(x_H,x_-)S_-(x_-) = \phi(x_H) = K(x_H,x_+)S_+(x_+) \quad . \tag{5.3}$$

So that, setting consideration of normalisation on one side, it might be tempting to write

$$S_-(x_-) = K^{-1}(x_H,x_-)K(x_H,x_+)S_+(x_+)$$

or, even,

$$S_+(x_+) = K^{-1}(x_H,x_+)K(x_H,x_-)S_-(x_-) \quad .$$

It is not really surprising that this does not work: simply because an amplitude can be the result of a range of source or sink values. The inverse notation will only mean anything within an integral over all the space-time values of $x_H$ . This, and the form of the 'inverse', are prefigured in the equation (2.8) of motion of $K(x,y)$ . Sadly, the integral that can be taken is over the hyperplane $H$ – one dimension short of a full measure.

Expanding both sides of equation 3 gives

$$\int d^4p' \, \frac{\gamma^\nu p'_\nu + mc}{p'^\nu p'_\nu - m^2c^2} e^{-ip'(x_H-x_-)/\hbar} S_-(x_-) = \int d^4p \, \frac{\gamma^\nu p_\nu + mc}{p^\nu p_\nu - m^2c^2} e^{-ip(x_H-x_+)/\hbar} S_+(x_+) \quad . \tag{5.4}$$

Assuming, for the moment, that the hyperplane $H$ is a constant-time surface, and integrating over the resulting 3-space, gives three $\delta$-functions on each side – leaving $\mathbf{p}' = \mathbf{0}$ and $\mathbf{p} = \mathbf{0}$ ; a natural consequence of the fact that $\mathbf{x}_+ = \mathbf{x}_-$ has been implicitly assumed. Thus equation 4 now reads

$$\int dp'_0 \, \frac{\gamma^0 p'_0 + mc}{(p'_0)^2 - m^2c^2} e^{-ip'_0(x^0_H - x^0_-)/\hbar} S_-(x_-) = \int dp \, \frac{\gamma^0 p_0 + mc}{p_0{}^2 - m^2c^2} e^{-ip_0(x^0_H - x^0_+)/\hbar} S_+(x_+) \quad .$$

This brings me back to the subject of contour integrals, as seems inevitable with propagators, though not to quite the horrors of my first formula for $K(x,y)$ . Concentrating on the integral:

$$\int \frac{z}{z^2 - M^2} e^{izt} \, dz = \int z e^{izt} \frac{1}{2M} \left( \frac{1}{z-M} - \frac{1}{z+M} \right) dz$$

The residues are:

$$\frac{Me^{iMt}}{2M} \quad \text{at} \quad z = M, \qquad \text{and} \qquad -\frac{-Me^{-iMt}}{2M} \quad \text{at} \quad z = -M.$$

Using a contour that includes both poles, as argued in the previous section, the integral is

$$\begin{aligned} \int \frac{z}{z^2 - M^2} e^{izt} \, dz &= 2\pi i \left( \frac{e^{iMt}}{2} + \frac{e^{-iMt}}{2} \right) \\ &= 2\pi i \cos Mt \quad . \end{aligned} \tag{5.5}$$

Likewise,

$$\int \frac{e^{izt}}{z^2 - M^2} \, dz = -2\pi \frac{\sin Mt}{M} \quad . \tag{5.6}$$

This means, since $x^0_H = (x^0_- + x^0_+)/2$ ,

$$\left[ i\gamma^0 \cos\left( mc \frac{x^0_- - x^0_+}{2\hbar} \right) - \sin\left( mc \frac{x^0_- - x^0_+}{2\hbar} \right) \right] S_-(x_-) =$$
$$\left[ i\gamma^0 \cos\left( mc \frac{x^0_+ - x^0_-}{2\hbar} \right) - \sin\left( mc \frac{x^0_+ - x^0_-}{2\hbar} \right) \right] S_+(x_+)$$

i.e., multiplying through by $-i$ ,

$$\left[ \gamma^0 \cos\left( mc \frac{x^0_- - x^0_+}{2\hbar} \right) + i \sin\left( mc \frac{x^0_- - x^0_+}{2\hbar} \right) \right] S_-(x_-) =$$
$$\left[ \gamma^0 \cos\left( mc \frac{x^0_+ - x^0_-}{2\hbar} \right) + i \sin\left( mc \frac{x^0_+ - x^0_-}{2\hbar} \right) \right] S_+(x_+)$$

Taking $\gamma^0$ to have its usual representation,

$$\gamma^0 = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad ,$$

then the trigonometric factors give exponential factors that are reciprocal. That is

$$S_-(x_-) = \begin{pmatrix} I_2 \exp\left(imc\frac{x_+^0 - x_-^0}{\hbar}\right) & 0_2 \\ 0_2 & I_2 \exp\left(imc\frac{x_-^0 - x_+^0}{\hbar}\right) \end{pmatrix} S_+(x_+) \tag{5.7}$$

– an expression of almost frightening simplicity, palliated by the severity of the assumption made about the hyperplane, $H$ .

If $H$ is now allowed to be any space-like hyperplane, so that $\mathbf{x}_+ \neq \mathbf{x}_-$ , there is a Lorentz boost that makes it a constant-time hyperplane. Writing this Lorentz transformation as

$$\begin{aligned} \mathbf{x}^* &= \alpha(\mathbf{x} - \boldsymbol{\beta} x^0) \\ x^{*0} &= \alpha(x^0 - \boldsymbol{\beta} \cdot \mathbf{x}) \end{aligned} \tag{5.8}$$

where $\alpha = (1 - \boldsymbol{\beta} \cdot \boldsymbol{\beta})^{-1/2}$ . Then, if it is required that $\mathbf{x}_+^* = \mathbf{x}_-^*$ (which is equivalent to $H$ being a surface of constant $x^{*0}$ ), this may be written as

$$\alpha(\mathbf{x}_+ - \boldsymbol{\beta} x_+^0) = \alpha(\mathbf{x}_- - \boldsymbol{\beta} x_-^0)$$

therefore,

$$\boldsymbol{\beta} = \frac{\mathbf{x}_+ - \mathbf{x}_-}{x_+^0 - x_-^0} \quad . \tag{5.9}$$

If the operator $\hat{U}(\boldsymbol{\beta})$ denotes this boost, the expansion of equation 4 can be integrated over the surface of constant $x^{*0}$, i.e., $H$ , as follows.

$$\int_{\mathbb{R}^4} d^4p' \, \hat{U}^{-1}(\boldsymbol{\beta}) \left[ \int d^3\mathbf{x}_H^* e^{-ip'^* x_H^*/\hbar} \hat{U}(\boldsymbol{\beta}) \left[ e^{ip'x_-/\hbar} \frac{\gamma^\nu p'_\nu + mc}{p'^\nu p'_\nu - m^2c^2} S_-(x_-) \right] \right] =$$

$$(2\pi\hbar)^3 \int_{\mathbb{R}^4} dp_0 \, \hat{U}^{-1}(\boldsymbol{\beta}) \left[ \int_{\mathbb{R}^3} d^3\mathfrak{p}^* e^{-ip_0^* x_H^{*0}/\hbar} \delta(\mathfrak{p}^*) \hat{U}(\boldsymbol{\beta}) \left[ e^{ipx_+/\hbar} \frac{\gamma^\nu p_\nu + mc}{p^\nu p_\nu - m^2c^2} S_+(x_+) \right] \right]$$

The integral over $\mathfrak{p}^*$ (likewise the one over $\mathfrak{p}^{*\prime}$ ) impliments a constraint:

$$\mathfrak{p}^* = 0 = \alpha(\mathfrak{p} + \boldsymbol{\beta} p_0) \quad ,$$

so that

$$\mathfrak{p} = -\boldsymbol{\beta} p_0 \quad ;$$

and, since

$$x_H^{*0} = \frac{x_-^{*0} + x_+^{*0}}{2} \quad ,$$

then

$$\begin{aligned} p_0^* x_H^{*0} &= \alpha(p_0 + \boldsymbol{\beta} \cdot \mathfrak{p}) \alpha \left( \frac{x_-^0 + x_+^0}{2} - \boldsymbol{\beta} \cdot \frac{\mathbf{x}_- + \mathbf{x}_+}{2} \right) \\ &= (1 - \boldsymbol{\beta} \cdot \boldsymbol{\beta})^{-1} (p_0 - \boldsymbol{\beta} \cdot \boldsymbol{\beta} p_0) \left( \frac{x_-^0 + x_+^0}{2} - \boldsymbol{\beta} \cdot \frac{\mathbf{x}_- + \mathbf{x}_+}{2} \right) \\ &= p_0 \left( \frac{x_-^0 + x_+^0}{2} - \boldsymbol{\beta} \cdot \frac{\mathbf{x}_- + \mathbf{x}_+}{2} \right) \quad . \end{aligned}$$

**Therefore**

$$\int_{\Re} dp' \exp\left[-\frac{ip'_0}{\hbar}\left(\frac{x^0_- + x^0_+}{2} - \boldsymbol{\beta}\cdot\frac{\mathbf{x}_- + \mathbf{x}_+}{2}\right)\right]\exp\left[\frac{ip'_0 x^0}{\hbar} + \frac{-ip'_0\boldsymbol{\beta}\cdot\mathbf{x}_-}{\hbar}\right]\times$$
$$\frac{\gamma^0 p'_0 + \boldsymbol{\gamma}\cdot(-\boldsymbol{\beta}p'_0) + mc}{(p'_0)^2 - \boldsymbol{\beta}\cdot\boldsymbol{\beta}(p'_0)^2 - m^2c^2} S_-(x_-) =$$
$$\int_{\Re} dp \exp\left[\frac{ip_0}{\hbar}\left(\frac{x^0_+ - x^0_-}{2} - \boldsymbol{\beta}\cdot\frac{\mathbf{x}_+ - \mathbf{x}_-}{2}\right)\right]\frac{(\gamma^0 - \boldsymbol{\gamma}\cdot\boldsymbol{\beta})p_0 + mc}{(1-\boldsymbol{\beta}\cdot\boldsymbol{\beta})(p_0)^2 - m^2c^2} S_+(x_+) \quad (5.10)$$

The poles are now at $p_0 = \pm\alpha mc$ , but otherwise these integrals are calculated exactly as in the previous, special case. Thus, applying equations 5 and 6 gives

$$\left[i\frac{\gamma^0 - \boldsymbol{\gamma}\cdot\boldsymbol{\beta}}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\cos\left(\frac{mc}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\frac{x^0_- - x^0_+ - \boldsymbol{\beta}\cdot(\mathbf{x}_- - \mathbf{x}_+)}{2\hbar}\right) - \right.$$
$$\left.\frac{mc}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\frac{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}{mc}\sin\left(\frac{mc}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\frac{x^0_- - x^0_+ - \boldsymbol{\beta}\cdot(\mathbf{x}_- - \mathbf{x}_+)}{2\hbar}\right)\right]S_-(x_-) =$$
$$\left[i\frac{\gamma^0 - \boldsymbol{\gamma}\cdot\boldsymbol{\beta}}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\cos\left(\frac{mc}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\frac{x^0_+ - x^0_- - \boldsymbol{\beta}\cdot(\mathbf{x}_+ - \mathbf{x}_-)}{2\hbar}\right) - \right.$$
$$\left.\frac{1}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\sin\left(\frac{mc}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\frac{x^0_+ - x^0_- - \boldsymbol{\beta}\cdot(\mathbf{x}_+ - \mathbf{x}_-)}{2\hbar}\right)\right]S_+(x_+)$$

Taking the $\gamma^\nu$ matrices to have their usual form:

$$\gamma^0 = \begin{pmatrix} I_2 & 0_2 \\ 0_2 & -I_2 \end{pmatrix}$$
$$\boldsymbol{\gamma}\cdot\boldsymbol{\beta} = \begin{pmatrix} 0_2 & \boldsymbol{\sigma}\cdot\boldsymbol{\beta} \\ -\boldsymbol{\sigma}\cdot\boldsymbol{\beta} & 0_2 \end{pmatrix}$$

where the $\sigma$ are the Pauli spin matrices.

Denoting the arguments of the trigonometric functions by

$$\pm\phi = \pm\frac{mc}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\frac{x^0_\pm - x^0_\mp - \boldsymbol{\beta}\cdot(\mathbf{x}_\pm - \mathbf{x}_\mp)}{2\hbar} \quad ,$$

the matrix equation becomes

$$\alpha^2\begin{pmatrix} (\cos\phi - i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 & -\boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi \\ \boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi & (-\cos\phi - i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 \end{pmatrix}S_-(x_-) =$$
$$\alpha^2\begin{pmatrix} (\cos\phi + i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 & -\boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi \\ \boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi & (-\cos\phi + i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 \end{pmatrix}S_+(x_+)$$

Now multiply through by

$$\begin{pmatrix} (\cos\phi + i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 & -\boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi \\ \boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi & (-\cos\phi + i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 \end{pmatrix}$$

On the left-hand side, the off-diagonal terms vanish, and because the Pauli spin matrices anti-commute the result is

$$\begin{pmatrix} \left(\frac{\cos^2\phi}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}} + \sin^2\phi - \frac{\boldsymbol{\beta}\cdot\boldsymbol{\beta}\cos^2\phi}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\right)I_2 & 0_2 \\ 0_2 & (\cos^2\phi + \sin^2\phi)I_2 \end{pmatrix}S_-(x_-) = \cdots$$

i.e.,

$$S_-(x_-) = \begin{pmatrix} (\cos\phi + i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 & -\boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi \\ \boldsymbol{\sigma}\cdot\boldsymbol{\beta}\cos\phi & (-\cos\phi + i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin\phi)I_2 \end{pmatrix} \frac{S_+(x_+)}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}$$

$$= \begin{pmatrix} \left((\cos^2\phi - \sin^2\phi)(1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}) + 2i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\cos\phi\sin\phi\right)I_2 & -(2i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\cos\phi\sin\phi)\boldsymbol{\sigma}\cdot\boldsymbol{\beta} \\ (2i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\cos\phi\sin\phi)\boldsymbol{\sigma}\cdot\boldsymbol{\beta} & \left((1-\boldsymbol{\beta}\cdot\boldsymbol{\beta})\cos 2\phi - i\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}\sin 2\phi\right)I_2 \end{pmatrix} \frac{S_+(x_+)}{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}$$

$$= \begin{pmatrix} \left(\cos 2\phi + \frac{i\sin 2\phi}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\right)I_2 & -\left(\frac{i\sin 2\phi}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\right)\boldsymbol{\sigma}\cdot\boldsymbol{\beta} \\ \left(\frac{i\sin 2\phi}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\right)\boldsymbol{\sigma}\cdot\boldsymbol{\beta} & \left(\cos 2\phi - \frac{i\sin 2\phi}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\right)I_2 \end{pmatrix} S_+(x_+)$$

i.e.,

$$S_-(x_-) = \left(\frac{i\sin 2\phi}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}(\gamma^0 - \boldsymbol{\beta}\cdot\boldsymbol{\gamma}) + \cos 2\phi\right) S_+(x_+)$$

Note that, using equation 9,

$$\begin{aligned} 2\phi &= \frac{mc}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}}\frac{x_+^0 - x_-^0 - \boldsymbol{\beta}\cdot(\mathbf{x}_+ - \mathbf{x}_-)}{\hbar} \\ &= \frac{mc(x_+^0 - x_-^0)}{\sqrt{(x_+^0 - x_-^0)^2 - |\mathbf{x}_+ - \mathbf{x}_-|^2}}\frac{(x_+^0 - x_-^0)^2 - |\mathbf{x}_+ - \mathbf{x}_-|^2}{(x_+^0 - x_-^0)\hbar} \\ &= \frac{mc}{\hbar}\sqrt{(x_+^0 - x_-^0)^2 - |\mathbf{x}_+ - \mathbf{x}_-|^2} \end{aligned} \tag{5.11}$$

Thus

$$S_-(x_-) = \left(\frac{i\sin\left(\frac{mc}{\hbar}\sqrt{(x_+^0 - x_-^0)^2 - |\mathbf{x}_+ - \mathbf{x}_-|^2}\right)\left(\gamma^0(x_+^0 - x_-^0) - \boldsymbol{\gamma}\cdot(\mathbf{x}_+ - \mathbf{x}_-)\right)}{\sqrt{(x_+^0 - x_-^0)^2 - |\mathbf{x}_+ - \mathbf{x}_-|^2}} + \cos\left(\frac{mc}{\hbar}\sqrt{(x_+^0 - x_-^0)^2 - |\mathbf{x}_+ - \mathbf{x}_-|^2}\right)\right) S_+(x_+) \tag{5.12}$$

– an expression of sterling invariance.

§§5.5 *Space-like separated sources and sinks*

Things are already somewhat awry if $(x_+)$ and $(x_-)$ are space-like separated: there is no solution for $\boldsymbol{\beta}$ if $x_-^0 = x_+^0$ , and otherwise $|\boldsymbol{\beta}| > 1$ .

Now, so far the only guarantee that this formalism conforms to the Special Principle of Relativity is a vague reference to the theory of characteristics. Well, rather than slog through the heavy-duty functional analysis involved in this, fascinating as that would be, a direct proof will be presented. Specifically, I will show that a source and a sink cannot be linked by way of an event, $(x_H)$ , that is space-like with respect to one or both.

With no loss of generality, I shall take $x_-^0 = x_+^0$ . The right-hand side of equation 4 now reads

$$\int d^4p\, e^{ip\cdot(x_+ - x_H)/\hbar}\frac{\gamma^\nu p_\nu + mc}{p^\nu p_\nu - m^2c^2}S_+(x_+) =$$

$$\int d^3\mathbf{p}\, e^{i\mathbf{p}\cdot(\mathbf{x}_+ - \mathbf{x}_H)/\hbar}\left(\gamma^0\int_{-\infty}^{\infty}\frac{p_0}{{p_0}^2 - (\mathbf{p}\cdot\mathbf{p} + m^2c^2)}dp_0 + (\boldsymbol{\gamma}\cdot\mathbf{p} + mc)\int_{-\infty}^{\infty}\frac{dp_0}{{p_0}^2 - (\mathbf{p}\cdot\mathbf{p} + m^2c^2)}\right)S_+(x_+)$$

The two integrals of interest in this show a near-diabolical familiarity; allowing them to be quickly dispatched:

$$\int_{-\infty}^{\infty} \frac{z}{z^2 - M^2} dz$$

– an odd integral over an even interval, and so zero;

$$\int_{-\infty}^{\infty} \frac{dz}{z^2 - M^2} = \frac{1}{2M} \int_{-\infty}^{\infty} \frac{1}{z - M} - \frac{1}{z + M}$$

$$= \frac{1}{2M} \times 2\pi i \times (1 - 1) = 0$$

It is therefore clear that $S_+(x_+)$ can influence nothing via events that are space-like separated from $(x_+)$ . This does not mean that $S_+(x_+)$ and $S_-(x_-)$ are unconnected – even if $(x_-)$ and $(x_+)$ are space-like separated – it is just that the linking event, $(x_H)$ , must lie either in their common past, or their common future: any two light-cones will eventually intersect.

Note that, as expected, Feynman's rule does not give this result – there is no cancellation in the second integral.

## §6 Probability, Energy, and all that

### §§6.1 *Probabilities: where they come from*

For a single point source leading to a single point sink, a simple consequence of equation 5.12 is that

$$\overline{S_-(x_-)} S_-(x_-) = \overline{S_+(x_-)} S_+(x_+) \tag{6.1}$$

where the adjoint spinor, $\overline{S}$ , is given by

$$\overline{S} = S^* \gamma^0$$

– $S^*$ is the transpose of the complex conjugate of $S$ . The presence of $\gamma^0$ means that this is an equation of quantities that are not positive-definite. I would like to ascribe a probability interpretation to this equation nonetheless. The interpretation is this:

$$\overline{S(x)} S(x) = pr_\epsilon - pr_\pi \qquad (pr_\epsilon, pr_\pi > 0)$$

$pr_\epsilon$ is the, so far unnormalised, probability (density) that there is an electron at $(x)$ : $pr_\pi$ is the same thing for a positron. This suggests that the spinor $S$ can be decomposed into two parts:

$$S = \begin{pmatrix} S_\epsilon \\ S_\pi \end{pmatrix} \quad . \tag{6.2}$$

$S_\epsilon$ a two component spinor representing the electron, and $S_\pi$ the corresponding representative of the positron.

The formula $\overline{S(x)} S(x) = S(x)^* \gamma^0 S(x)$ is an invariant, whence the two probability densities are also invariants. To see this, start from equation 5.1 again:

$$\left( i\hbar\gamma^\mu \frac{\partial}{\partial x^\mu} - mc \right) \psi(x) = S(x) \tag{6.3}$$

The Principle of Relativity demands that this formula be form-invariant, so that in another coordinate frame (denoted by primed quantities),

$$\left( i\hbar\gamma'^\mu \frac{\partial}{\partial x'^\mu} - mc \right) \psi'(x') = S'(x') \tag{6.4}$$

where $x'^{\nu} = \Lambda^{\nu}_{\mu} x^{\mu} + a^{\nu}$ . A lengthy but elementary proof, given with some clarity by R. H. Good,[10] shows that the $\gamma$–matrices are unique up to a unitary transformation; so nothing is lost by using the unprimed set throughout. Also

$$\frac{\partial}{\partial x^{\mu}} = \frac{\partial x'^{\nu}}{\partial x^{\mu}} \frac{\partial}{\partial x'^{\nu}} = \Lambda^{\nu}_{\mu} \frac{\partial}{\partial x'^{\nu}} \tag{6.5}$$

Further to the Principle of Relativity, there must be a transformation, $U(\Lambda, a)$ , that maps $\psi$ onto $\psi'$ :

$$\psi'(x') = U(\Lambda, a)\psi(x) \tag{6.6}$$

i.e.,

$$\psi(x) = U^{-1}(\Lambda, a)\psi'(x') \quad . \tag{6.7}$$

Using the substitutions of equations 5 and 7 in formula 3; and multiplying on the left by $U(\Lambda, a)$ , gives

$$\left( i\hbar U(\Lambda, a)\gamma^{\mu} U^{-1}(\Lambda, a)\Lambda^{\nu}_{\mu} \frac{\partial}{\partial x'^{\nu}} - mc \right) \psi'(x') = U(\Lambda, a) S(x) \tag{6.8}$$

The standard proof of the covariance of the Dirac equation (see, for example, Bjorken and Drell's book,[11] pp 18–25) next proceeds to find the representation of the Poincaré group so that the left-hand sides of equations 8 and 4 are identical, i.e., such that

$$\gamma^{\nu} = U(\Lambda, a)\gamma^{\mu} U^{-1}(\Lambda, a)\Lambda^{\nu}_{\mu}$$

or,

$$\Lambda^{\nu}_{\mu}\gamma^{\mu} = U^{-1}(\Lambda, a)\gamma^{\nu} U(\Lambda, a) \quad .$$

The representation is a faithful one of the Lorentz group, so that the translation group forms the kernel of the representation.

Finally, in order to complete this application of the Principle of Relativity, it is only necessary to equate the right-hand sides of formulae 8 and 4:

$$S'(x') = U(\Lambda, a) S(x) \tag{6.9}$$

– an equation identical in form to the transformation formula 6, as might have been expected.

One more result will be needed from Bjorken and Drell's book[11] (p 23, equation 2.26):

$$U^{-1}(\Lambda, a) = \gamma^0 \left(U(\Lambda, a)\right)^{\dagger} \gamma^0 \tag{6.10}$$

The invariance of $\overline{S}S$ can now be demonstrated:

$$\begin{aligned}
\overline{S}S &= S^{\dagger}(x)\gamma^0 S(x) \\
&= \left(U^{-1}(\Lambda, a)S'(x')\right)^{\dagger} \gamma^0 U^{-1}(\Lambda, a) S'(x') \\
&= \left(S'(x')\right)^{\dagger} \left(U^{-1}(\Lambda, a)\right)^{\dagger} \gamma^0 U^{-1}(\Lambda, a) S'(x') \\
&= \left(S'(x')\right)^{\dagger} \gamma^0 U(\Lambda, a) U^{-1}(\Lambda, a) S'(x') \\
&= \left(S'(x')\right)^{\dagger} \gamma^0 S'(x') = \overline{S'}S' \quad .
\end{aligned}$$

This is all very straightforward, but what if there are two events at which the source function is non-zero, or more? Writing out equation 5.12 for two source events contributing to a single sink event:

$$S_{1-}(x_-) = \left( \frac{i \sin \phi_1}{\sqrt{1 - \beta_1 \cdot \beta_1}} (\gamma^0 - \beta_1 \cdot \gamma) + \cos \phi_1 \right) S_+(x_1)$$

$$S_{2-}(x_-) = \left( \frac{i \sin \phi_2}{\sqrt{1 - \beta_2 \cdot \beta_2}} (\gamma^0 - \beta_2 \cdot \gamma) + \cos \phi_2 \right) S_+(x_2)$$

then the amplitude at the sink is (employing the usual superposition of amplitudes)

$$S_-(x_-) = S_{1-}(x_-) + S_{2-}(x_-) \quad .$$

**Whence, from the foregoing, the measure of probability is**

$$\overline{S_-(x_-)}S_-(x_-) = \overline{S_{1-}(x_-)}S_{1-}(x_-) + \overline{S_{2-}(x_-)}S_{2-}(x_-) + \overline{S_{1-}(x_-)}S_{2-}(x_-) + \overline{S_{2-}(x_-)}S_{1-}(x_-)$$

**i.e.,**

$$\begin{aligned}
&= |S_\epsilon(x_1)|^2 + |S_\epsilon(x_2)|^2 - (|S_\pi(x_1)|^2 + |S_\pi(x_2)|^2) + \\
&2\Re\left[\left(\cos\phi_1\cos\phi_2 + \frac{(1-\beta_1\cdot\beta_2)\sin\phi_1\sin\phi_2}{\sqrt{1-\beta_2\cdot\beta_2}\sqrt{1-\beta_1\cdot\beta_1}} + i\left[\frac{\sin\phi_1\cos\phi_2}{\sqrt{1-\beta_1\cdot\beta_1}} - \frac{\sin\phi_2\cos\phi_1}{\sqrt{1-\beta_2\cdot\beta_2}}\right]\right)\times\right. \\
&\qquad \left.(S_\epsilon^*(x_2)S_\epsilon(x_1) - S_\pi^*(x_2)S_\pi(x_1))\right] + \\
&2\Im\left[\left(\frac{\sin\phi_1\cos\phi_2\sigma\cdot\beta_1}{\sqrt{1-\beta_1\cdot\beta_1}} - \frac{\sin\phi_2\cos\phi_1\sigma\cdot\beta_2}{\sqrt{1-\beta_2\cdot\beta_2}}\right)(S_\epsilon^*(x_2)S_\pi(x_1) - S_\pi^*(x_2)S_\epsilon(x_1))\right] + \\
&2\Re\left[\frac{\sin\phi_1\sin\phi_2}{\sqrt{1-\beta_2\cdot\beta_2}\sqrt{1-\beta_1\cdot\beta_1}}(\sigma\cdot\beta_2 - \sigma\cdot\beta_1)(S_\epsilon^*(x_2)S_\pi(x_1) - S_\pi^*(x_2)S_\epsilon(x_1))\right]
\end{aligned}$$

This is always a real number, however it is not very well bounded.

Even if I set $S_\pi(x_1) = 0 = S_\pi(x_2)$ , the remainder of this expression is of indefinite sign and potentially quite large (in modulus):

$$\begin{aligned}
&|S_\epsilon(x_1)|^2 + |S_\epsilon(x_2)|^2 + \\
&2\Re\left[\left(\cos\phi_1\cos\phi_2 + \frac{(1-\beta_1\cdot\beta_2)\sin\phi_1\sin\phi_2}{\sqrt{1-\beta_2\cdot\beta_2}\sqrt{1-\beta_1\cdot\beta_1}} + i\left[\frac{\sin\phi_1\cos\phi_2}{\sqrt{1-\beta_1\cdot\beta_1}} - \frac{\sin\phi_2\cos\phi_1}{\sqrt{1-\beta_2\cdot\beta_2}}\right]\right)\times\right. \\
&\qquad \left.S_\epsilon^*(x_2)S_\epsilon(x_1)\right]
\end{aligned}$$

This not necessarily positive because, as $|\beta_k| \to 1$ ,

$$\frac{\sin\phi_k}{\sqrt{1-\beta_k\cdot\beta_k}} \to \frac{mc}{\hbar}(x_k^0 - x_-^0)$$
$$\frac{\beta_k\sin\phi_k}{\sqrt{1-\beta_k\cdot\beta_k}} \to \frac{mc}{\hbar}(\mathrm{x}_k - \mathrm{x}_-)$$

For an electron, $m = 9.11\times10^{-31}\mathrm{kg}$, and $c = 3\times10^8\mathrm{ms}^{-1}$. $\hbar = 1.05\times10^{-34}\mathrm{kg}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}$. so $mc/\hbar = 2.6\times10^{12}$. This is quite a few electrons, or, perhaps, quite a few positrons. If this is to mean anything it must be that particle–anti-particle pairs are being created as a result of the interference from these two sources. This seems reasonable to me since there is a direct relation between the 'classical kinetic energy' ( $\propto mc^2/\sqrt{1-\beta_k\cdot\beta_k}$ ) and the increase in particle numbers. Notice that this creation process is necessitated by the initial assumption about where and when the electrons might begin – the sources – and the event at which detection occurs – the sink. This assumption, for $|\beta_k| \sim 1$ , means that the electron, or part-electron, at $S_\pm(x_k)$ travels at a speed quite close to that of light in order to be detected. The 'speed' just mentioned is arrived at by the eminently empirical formula 5.9, and bears a suitably loose connection to the propagation of the electron amplitude.

Conversely, if there is very little separation between $(x_1)$ and $(x_2)$ , so that $\beta_1$ and $\beta_2$ are virtually parallel, then

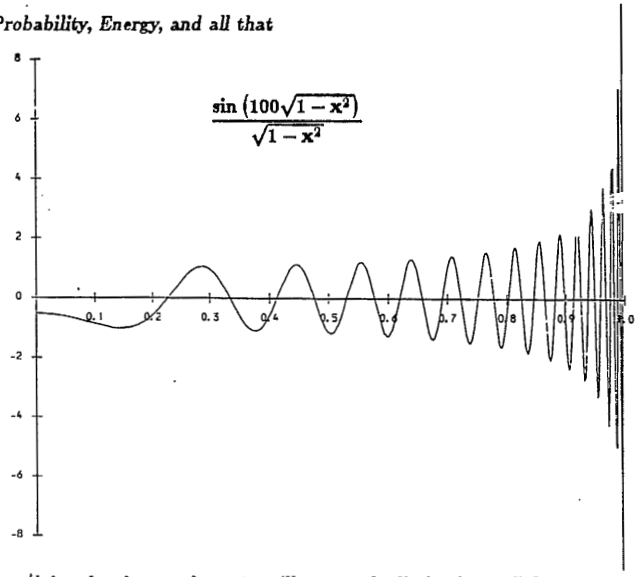
$$\phi_1 \approx \phi_2$$
$$1 - \beta_1\cdot\beta_2 \approx 1 - \beta_1\cdot\beta_1 \approx 1 - \beta_2\cdot\beta_2 \quad .$$

and the probability sum reduces to

$$\begin{aligned}
\overline{S_-}S_- \approx &|S_\epsilon(x_1)|^2 + |S_\epsilon(x_2)|^2 + 2\Re(S_\epsilon^*(x_2)S_\epsilon(x_1))\cos(\phi_1 - \phi_2) \\
&- (|S_\pi(x_1)|^2 + |S_\pi(x_2)|^2 + 2\Re(S_\pi^*(x_2)S_\pi(x_1))\cos(\phi_1 - \phi_2))
\end{aligned}$$

which will happily represent the free evolution of an electron or (and!) a positron in the low energy regime.

**The encouraging thing about the foregoing is that there are no infinities as long as all events are within some bounded volume of space-time. This may frustrate those theorists who like to use the word 'asymptotic', but seems otherwise a beneficial development.**




$mc/\hbar$ is taken here to be 100 to illustrate the limit of near-light speeds.

§§6.2 *Normalisation*

The expression, $\overline{S}S$ , is best interpreted as a probability rather than as a probability density. Therefore, separate normalisations must be applied to the source and sink. Thus, the source normalisation is obtained from the constraint

$$\int_{\mathbb{R}^4} \overline{S_+} S_+ \, d^4x = 1 \quad . \tag{6.11}$$

Having fixed the source, matters are more problematical for the detector, for which the obvious choice is to divide by the space-time volume occupied by the sink ( $\operatorname{supp} S_-$ ). The complications arise from the energy–momentum distribution and the creation and annihilation of particles. The motivation for this 'obvious choice' is straightforward, even if far from comprehensive: take a single point source at $(x_-)$ , then

$$\int \overline{S_-} S_- \, d^4x_- = \int_{\operatorname{supp} S_-} \overline{S_+(x_+)} S_+(x_+) \, d^4x_- = |\operatorname{supp} S_-| \quad .$$

If the source now occupies a block of space-time with dimensions greater than $\hbar/mc$ , and taking $S_\pi = 0$ and the low energy regime (i.e., put the sink within a low-velocity cone of the source), then the probability integral for a single point sink is

$$\overline{S_-} S_- = \int \int \overline{S_+(x_+)} S_+(x'_+) \, d^4x_+ \, d^4x'_+$$

Approximating the double integral by a double summation gives an expanded version of the formula in §§6.1

$$\sum_k |S_\epsilon(x_+)|^2 + \sum_l \sum_{k \neq l} 2\Re\left(S_\epsilon^*(x_k) S_\epsilon(x_l)\right) \cos(\phi_k - \phi_l)$$

Since the double summation is over a range of angles in excess of the wave length of the cosines, these terms can be counted on to cancel, mostly. The remaining 'diagonal sum' is the approximation of

$$\int \overline{S_+} S_+ \, d^4x_+ = 1 \quad .$$

Therefore, the weight of the source on each point of the sink is around unity, whence to normalise the sink the need to divide by its space-time volume. This normalisation process is meant to indicate that there was definitely a particle at the source and, subsequently, to give the probability that the particle was found at each given sink. It is, of course, possible that the same source function might produce non-zero amplitudes for sink functions elsewhere in space-time, but that, as the cliché goes, is another story.

### §§6.3 *Probabilities and Detectors*

The next point is going to make my remarks about constant-time hyperplanes appear faintly hypocritical, though I must claim that the order in which I have laid out this formalism means that I can now deduce a use for these surfaces rather than, as I have criticised, merely assuming their importance *ab initio*.

The thing is that detectors are spacial objects that function over periods of time. Thus, if the space-time support of a detector – during its active phase – is $\mathcal{D}$ then the spatial probability density it detects will be

$$f_\varepsilon(\mathbf{x}) - f_\pi(\mathbf{x}) = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \overline{S_-(x_-)}S_-(x_-)\,dx^0_- \tag{6.12}$$

A simple example is a photographic plate, which gathers on its emulsion an exposure dependent on time and intensity – not amplitude. Now in a boosted frame of reference the constant-time hyperplanes are different, and so too are such things as photographs. There is thus a place in an invariant theory for the coordinate–specific detectors generally in use.

In the high energy regime the large fluctuations in particle and anti-particle number appears to require a modification of this, perhaps taking into account the sensitivity of the detector, i.e., dividing $\mathcal{D}$ into intervals $\mathcal{D}_1, \mathcal{D}_2, \ldots$ and using

$$f(\mathbf{x}) = \frac{1}{|\mathcal{D}|}\sum_k \left| \int_{\mathcal{D}_k} \overline{S_-(x_-)}S_-(x_-)\,dx^0_- \right| \tag{6.13}$$

– since a photographic emulsion requires a certain exposure before it reacts, and such an exposure will register positrons and electrons in the same form.

A not dissimilar granularity must also be applied to the distribution $f(\mathbf{x})$ , since each particle in the photographic emulsion (or charge-coupled device, or whatever) registers the presence of a particle or particles, or their absence, for a volume of space. Further, the detector will only have a certain sensitivity, below which no distinction is made between a low probability and none at all. This is obviously not completely satisfactory. It is also not a subject to be pursued much further in what is intended to be a highly general and theoretical exposition. From the point of view of theory, the more relevant effect to analyse arises from the use of sources covering volumes of space.

A detector may, therefore, be effectively modelled as a sub-set of a constant-time hyperplane. The theoretical idealisation being made is that the intensities that the detector is adding up while it is active are roughly the same as those at a single instant. This will approximately be the case for low energy studies and highly sensitive detectors (e.g., fast shutter speeds).

As is fairly clear from the above, sources and sinks are primarily the means by which the preparation and detection of particles is transacted. In the next section intermediate sink–sources will be introduced, to represent what Feynman described as the scattering of the particle(s) by an external electromagnetic field. The creation and annihilation of particles is treated as a consequence of this formalism, rather than a foundation for it.

I am now in a position to elaborate on the provocative statements made all the way back at the beginning of §5 (*A justification of 'negative energies'*). For while some of the self-adjoint operators used in quantum theories serve as generators for various symmetries, to call them 'observables' and to claim that they all have a rigid connection with actual measurements is open to question. The positivist philosophy that inspired this convention is by now well criticised, if not completely discredited. The notion that only those numbers or marks appearing on measuring apparati have any reality (instrumentalism) is only of use to a philosopher: it was never a fruitful approach for physics. The encapsulation of the measurement process in a black box labelled 'momentum', or whatever, and then its representation by an operator is little short of a travesty of the delicate engineering required in experimental determinations. To imagine that theory need go no further than an operator is to ignore most of the physics of the quantum world – this has always been recognised in remarks about measurements disturbing the observed system.

The value of operators as generators of symmetries must not be underrated, but physical theory must look elsewhere, I maintain, for a model of measurements. The place to look, unsurprisingly, is inside the 'black box' measuring apparatus: if momentum is found by imposing a uniform magnetic field on some spatial volume then it is clear that any consistent quantum theory must do the same; a measurement of spin employs some form of Stern–Gerlach device – another imposition of an electromagnetic field; most measurements, in fact, seem to be about the spatial distribution arising from an arrangement of slits, electromagnetic fields, collimators and shutters: exactly the sorts of things that

it is hopeless to try to incorporate into Schrödinger's wave mechanics. If these simple components of reality can be adequately modelled there is a real chance that new insights may be gained into the microcosm.

It is my hope that the formalism outlined here will allow such real processes to be manageably modelled.

### §§8.4 *Energy, momentum, and spin*

Up to this point only the initial and final space-time distributions of a particle have been modelled. Now it is usual that the energy–momentum spectrum is also known. There must, therefore, be some way of expressing the notion that, though a high energy electron *could* get from $(x_+)$ to $(x_-)$ , this energy lies outside the spectrum of the *actual* electrons used.

If I abbreviate equation 5.12 to

$$S_-(x_-) = \mathrm{P}(m, x_-, x_+) S_+(x_+) \tag{6.14}$$

this is rather easily accomplished by using

$$S_-(x_-, p) = \mathrm{P}(m, x_-, x_+)\big(E(p) S_+(x_+)\big) \tag{6.15}$$

instead. $E$ is a probability amplitude function, $E : \mathfrak{R}^4 \to \mathcal{C}$ , reflecting the energy–momentum spectrum of the experimental particles, and it may even be taken to depend on $(x_+)$ if this is useful (two beams entering an experiment with differing spectra, for example).

The 4–vector $(p)$ is the empirical 4–momentum:

$$p^\nu = \frac{mc(x_-^\nu - x_+^\nu)}{\sqrt{(x_-^0 - x_+^0)^2 - |\mathbf{x}_- - \mathbf{x}_+|^2}}$$

Now in §3.1 of Bjorken and Drell's book[11] they show that a boost of the spinors

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\1\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

is (respectively)

$$u_{\epsilon\uparrow}(p) = \begin{pmatrix} \sqrt{\frac{p^0 + mc}{2mc}} \\ 0 \\ \frac{p^z}{\sqrt{2mc(p^0 + mc)}} \\ \frac{p^x + ip^y}{\sqrt{2mc(p^0 + mc)}} \end{pmatrix} \qquad u_{\epsilon\downarrow}(p) = \begin{pmatrix} 0 \\ \sqrt{\frac{p^0 + mc}{2mc}} \\ \frac{p^x - ip^y}{\sqrt{2mc(p^0 + mc)}} \\ \frac{-p^z}{\sqrt{2mc(p^0 + mc)}} \end{pmatrix}$$

$$u_{\pi\uparrow}(p) = \begin{pmatrix} \frac{p^z}{\sqrt{2mc(p^0 + mc)}} \\ \frac{p^x + ip^y}{\sqrt{2mc(p^0 + mc)}} \\ \sqrt{\frac{p^0 + mc}{2mc}} \\ 0 \end{pmatrix} \qquad u_{\pi\downarrow}(p) = \begin{pmatrix} \frac{p^x - ip^y}{\sqrt{2mc(p^0 + mc)}} \\ \frac{-p^z}{\sqrt{2mc(p^0 + mc)}} \\ 0 \\ \sqrt{\frac{p^0 + mc}{2mc}} \end{pmatrix}$$

Such that, for any $p$ ,

$$\overline{u_\sigma(p)} u_\rho(p) = \delta_{\sigma\rho}\omega_\sigma$$

$$\omega_\sigma = \begin{cases} +1 & , \text{ if } \sigma = \epsilon\uparrow , \text{ or } \epsilon\downarrow \\ -1 & , \text{ if } \sigma = \pi\uparrow , \text{ or } \pi\downarrow \end{cases} ;$$

**Significantly, this decomposition of a spinor is conserved by the modified propagator that I have proposed, viz.**

$$\mathrm{P}(m, x_-, x_+)u_\epsilon(p) = u_\epsilon(p)e^{i\phi}$$
$$\mathrm{P}(m, x_-, x_+)u_\pi(p) = u_\pi(p)e^{-i\phi}$$

**where, almost as before,**

$$\phi = \frac{mc}{\hbar}\sqrt{(x_-^0 - x_+^0)^2 - |\mathbf{x}_- - \mathbf{x}_+|^2}$$

**and**

$$\mathbf{p} = -\frac{\mathbf{x}_- - \mathbf{x}_+}{x_-^0 - x_+^0}p_0 \quad ,$$

**so that $(x_+)$ and $(x_-)$ are events linked by a world-line that, for a particle of rest mass, $m$, represents the classical trajectory with the 4–momentum, $(p)$ .**

The difference between a classical trajectory and the set of events that a spinor of definite 4–momentum being propagated through a sequence of events, is the factor of modulus one. In the limit of infinitesimal separation between the source and sink this difference between quantum and classical theories disappears.

An invariant decomposition of a source function in terms of spin. 4–position and 4–momentum is, therefore,

$$S_+(x,p) = s_{\epsilon\uparrow}(x,p)\,u_{\epsilon\uparrow}(p) + s_{\epsilon\downarrow}(x,p)\,u_{\epsilon\downarrow}(p) + s_{\pi\uparrow}(x,p)\,u_{\pi\uparrow}(p) + s_{\pi\downarrow}(x,p)\,u_{\pi\downarrow}(p) \qquad (6.16)$$

The four amplitude functions, $s_\mu$ , are derived by

$$s_\mu(x,p) = \langle u_\mu(p)|S_+(x,p) \qquad (6.17)$$

where $\langle u_\mu|$ is the projector onto the spinors spanned by $u_\mu$ (i.e., taking all four $u_\mu(p)$ as a basis).

It is now apparent that the source and sink functions are actually defined on phase space. This will not affect the previous discussions of normalisation if the momentum is integrated out – as has so far been implicitly assumed. This integral must take the measure

$$\frac{d^3\mathbf{p}}{\sqrt{\mathbf{p}\cdot\mathbf{p} + m^2c^2}}$$

to be invariant.

## §7 Applications and Illustrations

To wield a colourful metaphor at myself: it can be said that the rot sets in here. This part of the formalism is, really, a hopeful guess: the present section has some of the qualities of moonshine. The basis for checking the consistency and usefulness of this formalism will be presented, but no attempt will be made to pursue the issue further.

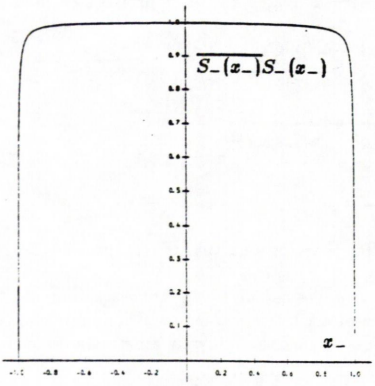
### §§7.1 *Some Exemplary Calculations*

The simplicity of formula 5.12 presents an immediate challenge: to calculate, even if rather roughly, some actual propagations. Following a variation on the usual quantum mechanical procedure (cf. the computation of the sink amplitude at a point arising from the source amplitudes at two source points on pp 19–20, above): a source function is defined on a number of points, then for any given point in the sink (i.e., choosing a $x_-$), formula 5.12gives the contribution to the sink amplitude from each source point. These contributions are added in exactly the same manner as the canonical superposition principle to give $S_-(x_-)$

$$S_-(x_-) = \sum_j \mathrm{P}(m, x_-, x_j)S_+(x_j)$$

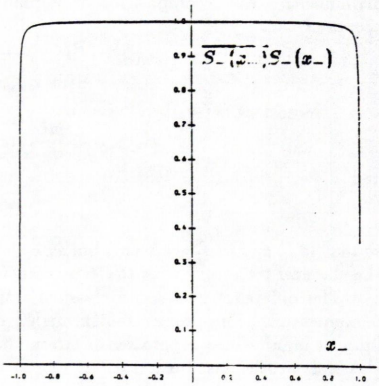
– where the source function is non-zero at the points $\{x_j\}$. Thence the probability of a particle appearing at $x_-$ is $\overline{S_-(x_-)}S_-(x_-)$.

This is illustrated in the next series of figures. The comparison between the first four shows that, at least in this case, only the centres of the sources need be modelled. For ease of computation I have taken $mc/\hbar = 100$ , and $x_-^0 - x_+^0 = 1$ .

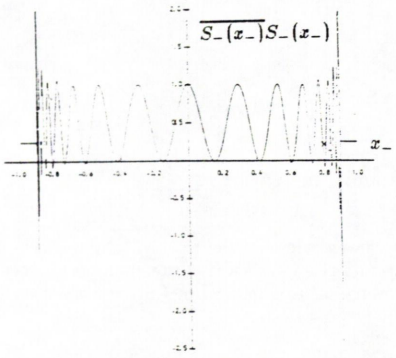



$S_+(x_j)=0.01$

$\{x_j=j\times10^{-5},\quad j\in\{1,\ldots,49,150,\ldots,199\}\}$

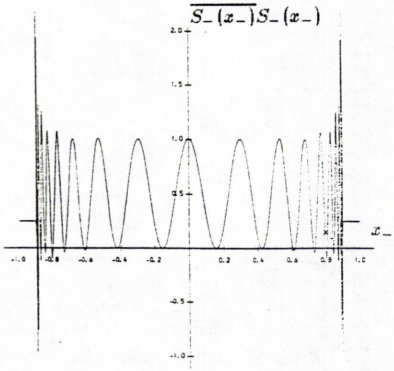



$S_+(0.0)=0.5\quad S_+(0.01)=0.5$

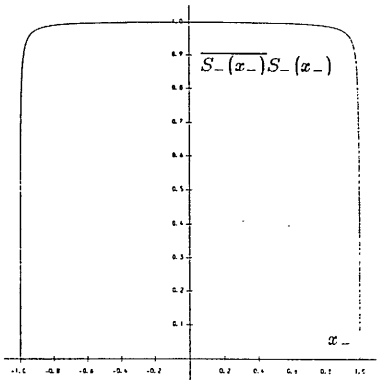



$S_+(x_j)=0.01$

$\{x_j=j\times10^{-5},\quad j\in\{-10099,\ldots,-10050,10000,\ldots10049\}\}$

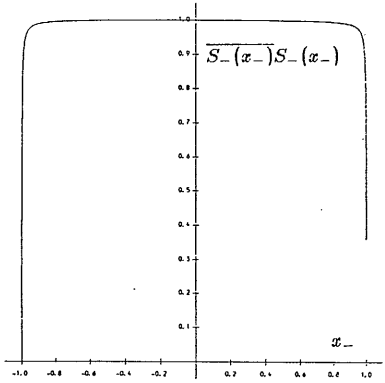



$S_+(-0.1)=0.5\quad S(0.1)=0.5$

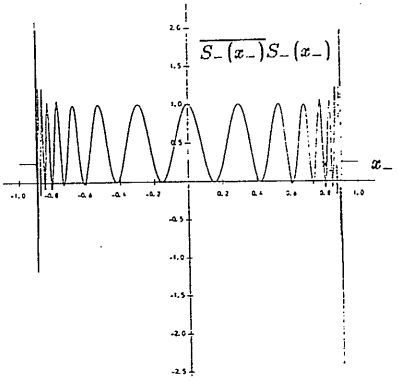



$S_+(x_j)=0.01$

$\{x_j=j\times10^{-5},\quad j\in\{1,\ldots,49,150,\ldots,199\}\}$

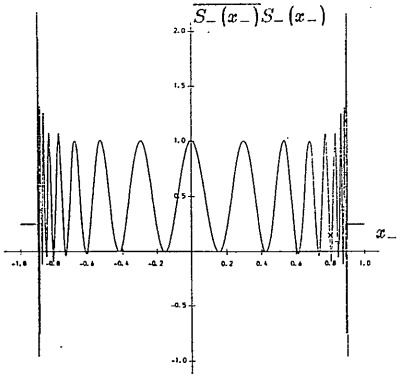



$S_+(0.0) = 0.5 \quad S_+(0.001) = 0.5$

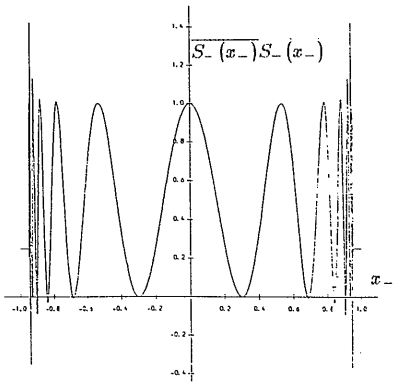



$S_+(x_j)=0.01$

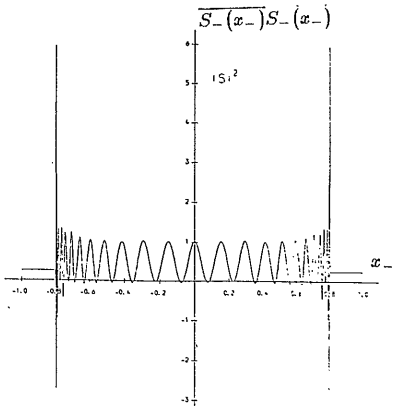
$\{x_j=j\times10^{-5},\quad j\in\{-10099,\ldots,-10050,10000,\ldots,10049\}\}$




$S_+(-0.1) = 0.5 \quad S_+(0.1) = 0.5$

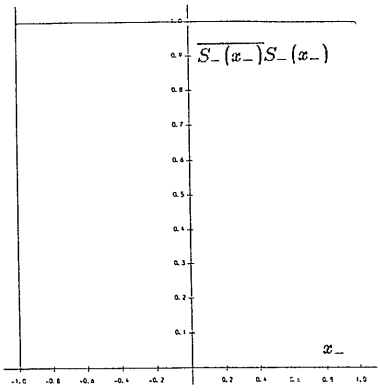



$S_+(-0.05) = 0.5 \quad S_+(0.05) = 0.5$

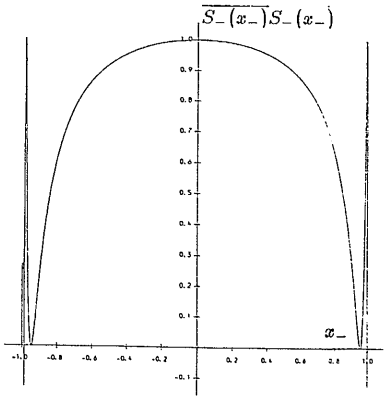



$S_+(-0.2) = 0.5 \quad S_+(0.2) = 0.5$

Using this considerable simplification the next four figures show the effect of varying the source separation. These computations only employ a single spatial axis, and for this reason rather over-simplify what might otherwise be a model of two-slit interference (of beams with unbounded momentum spectra).



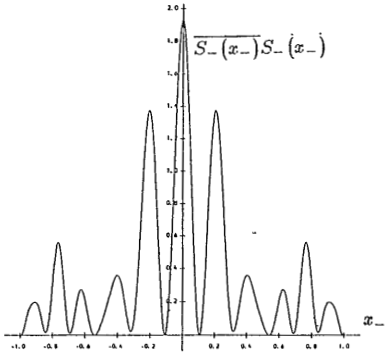

$S_+(0.0) = 0.5 \quad S_+(0.0001) = 0.5$



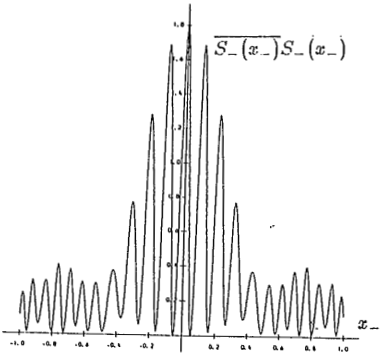

$S_+(0.0) = 0.5 \quad S_+(0.01) = 0.5$

The remaining three figures are the result of introducing a fixed distance in a second spatial direction between sources and sinks, together with an approximation of continuous output from the sources – by extending the source in time.

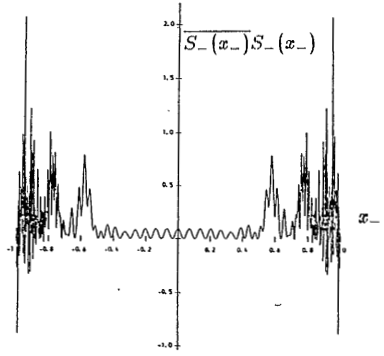
In these last examples I have set $mc/\hbar = 500$ ; there are only two spatial positions for the sources ( $(x_1, y)$ and $(x_2, y)$ ), but for each source-position is active for a range of times ( $\delta t_j$ ). The difference in $y$–values is fixed at $\sqrt{3}$ ; the amplitude at al the source points is taken as 0.127.




$x_1=-0.05, \quad x_2=0.05;$
$\delta t \in \{2.400,\ldots,2.455\}$




$x_1=-0.1, \quad x_2=0.1;$
$\delta t \in \{2.400,\ldots,2.455\}$




$x_1=-0.1, \quad x_2=0.1;$
$\delta t \in \{2.000,\ldots,2.055\}$

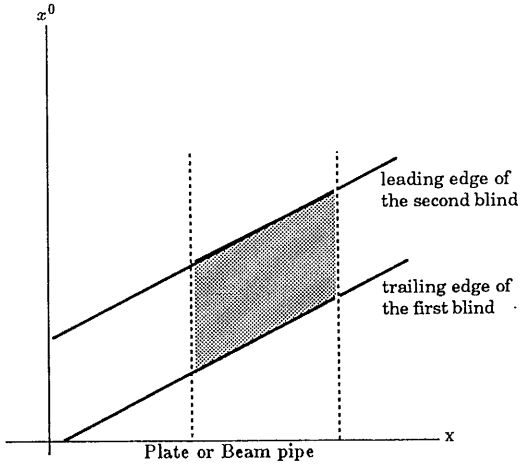
It should be noted that by dint of prescribing only the sources and sinks a considerable measure of control can be exercised over the momentum spectrum of the model.

§§7.2 *Electronic snap-shots; complementarity abandoned*

The two simplest devices to be modelled, and it is a curious fact that the model is the same, are the photographic plate and the end of a beam pipe. The difference between these is that one is a 'sink' and the other is a 'source' of particles.

In both cases there is a shutter. This may be taken to be a pair of moving blinds, as commonly found in SLR cameras. The edge of each blind is assumed to move at a constant velocity, leaving a slot through which particles can pass. For the photographic plate the shutter is additionally assumed to move immediately in front of the emulsion (a focal plane shutter); for the beam pipe the shutter is merely mounted on the end.

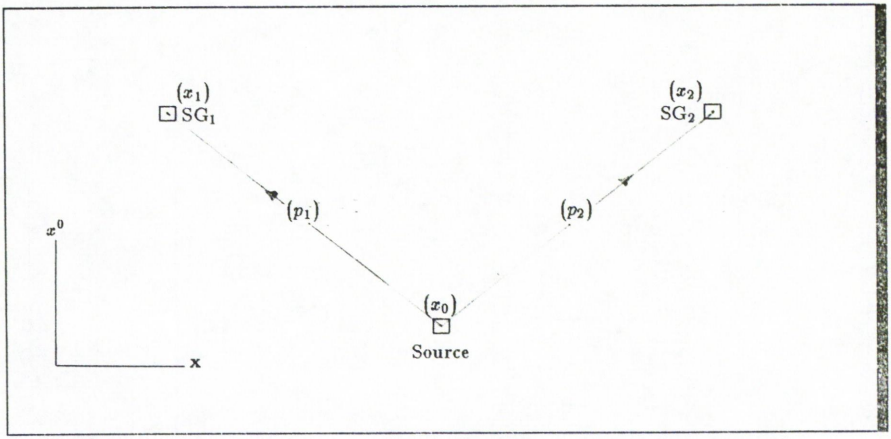
If the shutter lies in the $x$-$z$–plane, and the slot moves along the $x$–axis, then the following 'aerial view' fulfils the rôle of an adequate model:




Armed with this representation, it is possible to tackle Bohr's thesis of Complementarity as it is manifested by the so-called 'wave–particle duality'. By adopting the heretical notion that a physical theory is only a constrained description – a theory up to the degree of constraint imposed on the system – then the need for the dialectical nonsense of complementarity is avoided.

The free evolution of a system is described in a manner somewhat akin to that of a wave; any particle-like behaviour arises now from a high degree of spatial confinement. Thus, if the probability amplitude is calculated for an entire photographic plate the result is an interference pattern. If only a small area of the same plate is considered (i.e., forms the support of the sink) then the amplitude obtained represents the formation of a small blob at that specific point on the plate. There is no logical or physical incompatibility at work here – just as might have been suspected all along.

There remains the notion of complementary measurements, that is, pairs of measurements which, in the usual parlance, do not commute. If the actual processes of measurement are now to be modelled this becomes a trivial consequence of the facts of those processes: the spin cannot be measured in two different directions because this poses contradictory requirements on the electromagnetic field of the Stern-Gerlach device. Bell's theorem therefore involves expectation values of a set of quite distinct experiments – the paradox is resolved by avoiding comparisons that make no sense, not by abandoning locality. In the present context, the original paradox of Einstein, Podolsky and Rosen admits of a completely 'deterministic' solution. Though the theory proposed here is not necessarily complete, in the sense of Einstein *et al.*,; but, as the work of Karl Popper and Kurt Gödel suggests, completeness is a chimera.




The space-time layout of the EPR experiment

§§7.3 *Bohm's revision of the Einstein–Podolsky–Rosen thought experiment*

One major omission from the present theory is any mention of Pauli's Exclusion Principle. Whether the source normalisation can merely be changed, or whether some form of direct product of single particle representations must be devised, is an entirely open question. A simple example in which to explore this issue is the Einstein–Podolsky–Rosen thought experiment. The interaction of an incident beam with two Stern–Gerlach apparati and two cameras is just a specific application of the ideas of the previous sub-sections.

The simplest representation of the experiment uses point-like sources and sinks, and assumes the behaviour of the Stern–Gerlach apparati, $SG_i$ , (much as has usually been the case when an expectation value, $\langle\hat{\sigma}\rangle$ , was supposed to represent a measurement) – the issue of the exact functioning of the device that distinguishes spin orientations is, of course, merely postponed.

Source function: (fully symmetrised)

$$S_+(x_0,x_0) = \frac{1}{2}\big(u_{\epsilon\uparrow}(p_1)\otimes u_{\epsilon\downarrow}(p_2) - u_{\epsilon\downarrow}(p_1)\otimes u_{\epsilon\uparrow}(p_2) \\ -u_{\epsilon\downarrow}(p_2)\otimes u_{\epsilon\uparrow}(p_1) + u_{\epsilon\uparrow}(p_2)\otimes u_{\epsilon\downarrow}(p_1)\big)$$

Sink function:

$$S_-(x_1,x_2) = \frac{1}{2}\big(u_{\epsilon\uparrow}(p_1)\otimes u_{\epsilon\downarrow}(p_2) - u_{\epsilon\downarrow}(p_1)\otimes u_{\epsilon\uparrow}(p_2)\big)e^{i\phi_1}e^{i\phi_2}$$
$$S_-(x_2,x_1) = \frac{1}{2}\big(-u_{\epsilon\downarrow}(p_2)\otimes u_{\epsilon\uparrow}(p_1) + u_{\epsilon\uparrow}(p_2)\otimes u_{\epsilon\downarrow}(p_1)\big)e^{i\phi_1}e^{i\phi_2}$$

Strictly, the next step is to propagate $S_-(x_1,x_2)$ and $S_-(x_2,x_1)$ onto the events corresponding to the detection of the differing spin orientations being measured ( $x_{1\uparrow}$ , $x_{1\downarrow}$ , $x_{2\uparrow}$ , $x_{2\downarrow}$ , say). Bohm's correlation function can be found without recourse to this by decomposing the intermediate sink-sources in terms of the spins being measured:

$$u_{\epsilon\uparrow} = a_{1\uparrow}u_{1\uparrow} + a_{1\downarrow}u_{1\downarrow} = a_{2\uparrow}u_{2\uparrow} + a_{2\downarrow}u_{2\downarrow}$$
$$u_{\epsilon\downarrow} = b_{1\uparrow}u_{1\uparrow} + b_{1\downarrow}u_{1\downarrow} = b_{2\uparrow}u_{2\uparrow} + b_{2\downarrow}u_{2\downarrow} \quad .$$

Using just

$$S_-(x_1,x_2) = \frac{1}{\sqrt{2}}\big(u_{\epsilon\uparrow}(p_1)\otimes u_{\epsilon\downarrow}(p_2) - u_{\epsilon\downarrow}(p_1)\otimes u_{\epsilon\uparrow}(p_2)\big)$$

since the fully symmetrised treatment produces two copies of all terms and then divides by a half; and the $e^{i\phi}$'s add nothing:

$$S_+(x_1,x_2) = \frac{1}{\sqrt{2}}\big((a_{1\uparrow}u_{1\uparrow}+a_{1\downarrow}u_{1\downarrow})\otimes(b_{2\uparrow}u_{2\uparrow}+b_{2\downarrow}u_{2\downarrow}) \\ -(b_{1\uparrow}u_{1\uparrow}+b_{1\downarrow}u_{1\downarrow})\otimes(a_{2\uparrow}u_{2\uparrow}+a_{2\downarrow}u_{2\downarrow})\big)$$

Then

$$S_-(x_{1\uparrow},x_{2\uparrow}) = \frac{1}{\sqrt{2}}(a_{1\uparrow}u_{1\uparrow}\otimes b_{2\uparrow}u_{2\uparrow} - b_{1\uparrow}u_{1\uparrow}\otimes a_{2\uparrow}u_{2\uparrow})$$
$$S_-(x_{1\uparrow},x_{2\downarrow}) = \frac{1}{\sqrt{2}}(a_{1\uparrow}b_{2\downarrow} - b_{1\uparrow}a_{2\downarrow})u_{1\uparrow}\otimes u_{2\downarrow}$$
$$S_-(x_{1\downarrow},x_{2\uparrow}) = \frac{1}{\sqrt{2}}(a_{1\downarrow}b_{2\uparrow} - b_{1\downarrow}a_{2\uparrow})u_{1\downarrow}\otimes u_{2\uparrow}$$
$$S_-(x_{1\downarrow},x_{2\downarrow}) = \frac{1}{\sqrt{2}}(a_{1\downarrow}b_{2\downarrow} - b_{1\downarrow}a_{2\downarrow})u_{1\downarrow}\otimes u_{2\downarrow}$$

The correlation function is now found by weighting the probabilities at the four detecting events by the product of the spin 'eigen-values':

$$(1\times 1)\overline{S_-(x_{1\uparrow},x_{2\uparrow})}S_-(x_{1\uparrow},x_{2\uparrow}) + (1\times -1)\overline{S_-(x_{1\uparrow},x_{2\downarrow})}S_-(x_{1\uparrow},x_{2\downarrow}) \\ +(-1\times 1)\overline{S_-(x_{1\downarrow},x_{2\uparrow})}S_-(x_{1\downarrow},x_{2\uparrow}) + (-1\times -1)\overline{S_-(x_{1\downarrow},x_{2\downarrow})}S_-(x_{1\downarrow},x_{2\downarrow})$$

Which can be written in an altogether more conventional form as

$$\overline{S_-(x_1,x_2)}\,\dot{o}\,S_-(x_1,x_2) = -\cos(\theta_1-\theta_2) \quad .$$

This is perhaps not the briefest summary of a well-known result. I hope, in presenting it thus, to have illustrated the revised quantum mechanics I have developed, as well as the way that 'observables', where these are truly observable, are only a superficial mechanism to avoid more complete physical analysis.

Note that there is no need to invoke 'retrocausation' or any such causal influence that travels, first, backwards in time from a measurement with a definite outcome, and then forwards in time to dictate the outcome of a second measurement. Nor is there any 'collapse of the wave packet'. It is possible to produce two classes of model, however:

(i) The source at $(x_0)$ is deduced on the grounds of symmetry and conservation laws (e.g., the singlet or triplet state).

(ii) The source at $(x_0)$ is chosen so that the result of one measurement is for one orientation to have probability 1.

i and ii may be related, or even deduced from each other.

§§7.4 *Quantum Electrodynamics, well, maybe*

Leaving the comparatively safe pastures of quantum mechanics, I can now try setting the electromagnetic 4-potential to some non-zero value. Because the wave equation is now inhomogeneous, Feynman's perturbation expansion can no longer be deduced. The formulae governing the scattering of the probability amplitude may be guessed to be the same as Feynman's, since the justification in terms of scattering at the field points carries over – the field point at which scattering occurs, $(x_S)$ , becomes a sink, $S_-(x_S)$ , and then a source, $S_+(x_S)$ . The scattering process is incisively expressed:

$$S_+(x_S) = \frac{-ie}{c\hbar}\gamma^\nu A_\nu(x_S)S_-(x_S) \qquad (7.1)$$

The calculation of the amplitude at the detector (the final sink) is then the sum over the amplitudes contributed by all the sources: both the original source (zeroth order scattering) and all the sink-sources at field points. There is, as in Feynman's formulation, the possibility of an infinite number of scatterings, depending on how many times an $S_+(x_S)$ is used to contribute to the amplitude at any other field point, $(x'_S)$ say.

However, this will not do. The dimensions of $eA/c$ are those of momentum – whence the appearance of this in Dirac's equation – and therefore the dimensions of $eA/c\hbar$ are those of $(\text{length})^{-1}$.

This was perfect for Feynman's treatment because his amplitudes gave rise to a purely spatial probability density, which means that the perturbation expansion produces amplitudes that are again the precursors of spatial probability densities. For example, Feynman's first-order scattering correction is

$$-i\int K_+(2,3)\frac{e\not{A}(3)}{c\hbar}K_+(3,1)\psi\, d^4x_3$$

(in roughly Feynman's notation). The dimensions of the integrand are

$$\begin{aligned}\frac{e\not{A}(3)}{c\hbar} \quad &\cdots \quad (\text{length})^{-1}\\ K_+(3,1)\psi \quad &\cdots \quad (\text{length})^{-3/2}\\ d^4x_3 \quad &\cdots \quad (\text{length})^4\end{aligned}$$

so that, finally, $K_+(2,3)$ produces a spatial amplitude of dimensions, $(\text{length})^{-3/2}$ .

If a scattering scheme similar to Feynman's is to work for the present investigation then the factor multiplying the intermediate sink amplitude must be a dimensionless invariant. One candidate is

$$\frac{e\gamma^\nu A_\nu(x_S)}{mc^2}$$

or, marginally different,

$$\frac{e\gamma^\nu}{mc^2}\left(A_\nu(x_S)-A_\nu(x_+)\right) \quad .$$

Tempting as this might be, it can not be the whole story; since the propagation of amplitudes is constrained by the empirical energy–momentum spectrum, and scattering by an applied electromagnetic field changes this spectrum. The usual way of converting a theory for free particles to one in an external electromagnetic field is the substitution

$$p \rightarrow p - \frac{e}{c}A$$

for an electron, and

$$p \rightarrow p + \frac{e}{c}A$$

for a positron. This is because the momentum, $(P)$ , conjugate to position in the Hamiltonian form of classical electrodynamics (of a particle with charge, $q$ ) differs in this way from the kinematic or, as I have called it, the empirical momentum, $(p)$ :

$$P_\nu = p_\nu + \frac{q}{c}A_\nu \quad .$$

Thus, when the canonical quantisation procedure is applied to the invariant equation

$$\left(P-\frac{e}{c}A\right)^\nu\left(P-\frac{e}{c}A\right)_\nu = m^2c^2$$

it is, in fact, the equation

$$p^\nu p_\nu = m^2c^2$$

once again that is being 'quantised'. Since the approach I have adopted here uses only the empirical momentum, not the conjugate one, I can continue to use the same Dirac equation. It becomes necessary, however, to take into account the variation of the empirical momentum ( $dp$ ) that each propagation ( $x \rightarrow x + dx$ ) entrains.

The process of 'scattering by the field' is, therefore, reduced to the calculation of $dp$ for each component of a spinor – since this must be how a Stern–Gerlach apparatus functions. To begin with, spin will not be modelled. The action integral for an electron in an external field is[12]

$$W = \int mc\sqrt{dx^\nu\, dx_\nu} + \frac{e}{c}A_\nu\, dx^\nu \tag{7.2}$$

Taking a variation of this gives

$$\delta W = \int mu_\nu\, d\delta x^\nu + \frac{e}{c}A_\nu\, d\delta x^\nu + \frac{e}{c}\frac{\partial A_\nu}{\partial x^\mu}\,\delta x^\mu\, dx^\nu$$

Since: $c^2\,d\tau^2 = dx^\nu\,dx_\nu$ , so

$$2c\,d\tau\,\delta d\tau = 2\,dx_\nu\,d\delta x^\nu$$

thus
$$\delta(c\,d\tau) = \frac{\partial x^\nu}{\partial \tau}\,d\delta x^\nu = u_\nu\,d\delta x^\nu \quad ;$$

and

$$\delta A_\nu = \frac{\partial A_\nu}{\partial x^\mu}\,\delta x^\mu \quad .$$

Integrating by parts,

$$\delta W = -\int \delta x^\nu\,d\left(p_\nu - \frac{e}{c}A_\nu\right) + \int \frac{e}{c}\frac{\partial A_\mu}{\partial x^\nu}\,\delta x^\nu\,dx^\mu \quad .$$

If $\delta W = 0$ , and since $\delta x^\nu$ is arbitrary,

$$d\left(p_\nu - \frac{e}{c}A_\nu\right) = \frac{e}{c}\frac{\partial A_\mu}{\partial x^\nu}\,dx^\mu$$

or

$$dp_\nu = \frac{e}{c}\left(\frac{\partial A_\mu}{\partial x^\nu} - \frac{\partial A_\nu}{\partial x^\mu}\right)dx^\mu \quad . \tag{7.3}$$

I shall take $dx^\nu$ and $dp_\nu$ to be finite steps, so that the 'classical' results are recovered in the limit $dx^\nu \to 0$ of the 'quantum' theory.

Writing

$$F_{\mu\nu} = \frac{e}{c}\left(\frac{\partial A_\mu}{\partial x^\nu} - \frac{\partial A_\nu}{\partial x^\mu}\right)$$

one possible way of incorporating a dependence on spin into the expression for $dp$ is to add a new term to the action integral:

$$\int s^\mu(\sigma)F_{\mu\nu}\,dx^\nu$$

which relies on the existence of a 4–vector, $(s)$ , dependent on spin:

$$\sigma \in \{\epsilon\uparrow,\quad \epsilon\downarrow,\quad \pi\uparrow,\quad \pi\downarrow\} \quad .$$

This modifies the expression for $dp$ :

$$dp_\nu(\sigma) = \omega_\sigma F_{\mu\nu}\,dx^\mu + s^\rho(\sigma)\left(\frac{\partial F_{\rho\mu}}{\partial x^\nu} - \frac{\partial F_{\rho\nu}}{\partial x^\mu}\right)dx^\mu \quad . \tag{7.4}$$

§§7.5 *The double slit experiment*

This is an experiment involving, for my purposes, three parts: a source, an impenetrable plate bearing two slits, and a sink. The source may be taken to be the end of a beam pipe; the sink can be a simple camera. The new object is the slitted plate. Now an intermediate sink function can be defined to cover the whole plate, thus allowing the incident particle amplitude to be gauged. But it is only at the two slits that any particles progress further. To implement this the sink at the plate is multiplied by a function, $c(\mathbf{x}_P)$ , which represents the transmission cross-section at each $\mathbf{x}_P$ . In its simplest form this is

$$c(\mathbf{x}_P) = \begin{cases} 1 & \text{, if } \mathbf{x}_P \text{ lies in a slit;} \\ 0 & \text{, otherwise.} \end{cases} \tag{7.5}$$

The intermediate sink then acts as a source according to

$$S_+(x_P) = c(x_P)S_-(x_P) \quad . \tag{7.6}$$

Here, unlike in the previous sub-section, the original source does not contribute to the amplitude at the final sink (the camera). The final distribution is generated solely by the sources at the two slits. The sinks at the slits will act (i.e., have temporal extension) for exactly long enough for the camera to take its photograph.

Another difference between this sub-section and its predecessor is in the way the empirical 4–momentum is propagated. If amplitudes retained their 4–momentum when they pass through the slits then there would be no diffraction, and so, not much of an interference pattern. This is analogous to geometric optics. Obviously there must be some mechanism by which the various momenta are combined to produce some sort of diffraction. The simplest remedy would be to ignore the momentum probability distribution – by integrating it out.

Alternatively, if there are amplitudes

$$S_+(x_P, p_1) \qquad \text{and} \qquad S_+(x_P, p_2)$$

then the diffraction could simply be the non-linear combination of these:

$$S_+(x_P, p_1 + p_2) = S_+(x_P, p_1) \times S_+(x_P, p_2) \tag{7.7}$$

$$S_+(x_P, 2p_1 + p_2) = S_+(x_P, p_1) \times S_+(x_P, p_1) \times S_+(x_P, p_2) \tag{7.8}$$

where the products on the right are obtained by some form of (invariant) component–by–component multiplication, e.g.,

$$\begin{aligned} S_+(x_P, p_1) \times S_+(x_P, p_2) = {} & s_{\epsilon\uparrow}(x, p_1)\, s_{\epsilon\uparrow}(x, p_2)\, u_{\epsilon\uparrow}(p_1 + p_2) + s_{\epsilon\downarrow}(x, p_1)\, s_{\epsilon\downarrow}(x, p_2)\, u_{\epsilon\downarrow}(p_1 + p_2) + \\ & s_{\pi\uparrow}(x, p_1)\, s_{\pi\uparrow}(x, p_2)\, u_{\pi\uparrow}(p_1 + p_2) + s_{\pi\downarrow}(x, p_1)\, s_{\pi\downarrow}(x, p_2)\, u_{\pi\downarrow}(p_1 + p_2) \end{aligned}$$

Altering the source function like this is liable to change the normalisation, or, rather, require a modified normalisation procedure. Moreover, it appears that, having provided a general solution to Dirac's wave equation, it becomes necessary to contrive a new wave mechanism to deal with diffraction.

§§7.6 *A last word*

The theory sketched here is qualitatively different from the usual approach to quantum theories: it is descriptive rather than predictive. Given the space-time arrangement of an experiment, the propagation of electrons and/or positrons is described. In only a limited sense is any progression of states evolved from an initial state. Abandoning evolution is necessitated by the imposition of Poincaré–invariance; but this has the added advantage of removing the need for such convolutions as Wheeler's delayed-choice experiments, since the experiment performed and its result do not depend on when the choice is made – provided a choice is made. There is no elusive superluminal signalling to worry about.

Bohm's revision of the Einstein–Podolsky–Rosen experiment ceases to be paradoxical because Bell's theorem is a comparison of the statistics of a number of different, though similar, experiments. The original EPR thought-experiment is not at all mysterious in the present context.

The Problem of Locality ceases to be a problem the moment it is realised that operators may correspond to symmetries but they do not relate to real measurements. The absence of a position operator (or, which is the same thing, a multiplicity of such operators) is simply a consequence of the fact that no sub-group of the Poincaré group is generated by position. This was always a somewhat odd little mystery, as somewhere near the beginning of each formulation a four–dimensional space is introduced that can only meaningfully be Minkowski space-time. So that, in producing conundrums about position and locality, these concepts have already been established fairly unambiguously.

All told, I feel confident that this outline has more promise than previous assaults on the bastions of a 'relativistic quantum theory'. This is not least because there is a clear connection with that strange realm beyond theoretical ruminations where actual particles do their best to confound. There are, of course, many outstanding problems left to tackle.

## §8 Bibliography

(References are preceded by the page number on which they first occur.)

The various generalised functions used as propagators were first deduced by Dirac and Pauli:

W. Pauli, *Relativistic Field Theories Of Elementary Particles*, Rev. Mod. Phys. **13** (1941) 203.

P. A. M. Dirac, *Discussion of the infinite distribution of electrons in the theory of positrons*, Proc. Camb. Philos. Soc. **30** (1934) 150–163.

2: [1] Four papers make up Feynman's prize-winning Q.E.D. theory:

R. P. Feynman, *The Theory Of Positrons*, Phys. Rev. **76** (1949) 749–759.

R. P. Feynman, *Space-Time Approach To Quantum Electrodynamics*, Phys. Rev. **76** (1949) 769–789.

R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (1950) 440–457.

R. P. Feynman, *An Operator Calculus Having Applications In Quantum Electrodynamics*, Phys. Rev. **84** (1951) 108–128.

2: [2] S. S. Schweber, *Feynman and the visualization of space-time processes*, Rev. Mod. Phys. **58** (1986) 449–508.

2: [3] L. Gårding, *The mathematics of invariant wave equations*, pp 102–142 in *Invariant Wave Equations (Proceedings of a conference at Erice, 1977)*, Springer–Verlag Lecture Notes in Physics no. 73, 1978.

2: [4] D. Zwanziger, *Method of Characteristics in the External Field Problem*, or, *How to recognise an acausal equation when you see one*, pp 143–145 ibid.

2: [5] L. Hörmander, *Linear Partial Differential Operators*, Springer–Verlag, 1963.

2: [6] See Chapter 5.

3: [7] An excellent account of the theory of Green's functions and generalised functions can be found in

D. H. Griffel, *Applied Functional Analysis*, Ellis Horwood, 1981.

13: [8] Karl Popper has rigorously shown that no system can predict its own activity (*The Open Universe; an argument for indeterminism*, Hutchinson, 1982; pp 68–77) using a variation on the Stopping Problem for Turing machines. The idea of predicting the activity or evolution of a quantum system step–by–step as it happens is nearly as absurd. There is no value to be found in retaining this relic of a discredited interpretation of problem solving in classical mechanics. On the other hand, just because a theory does not purport to predict a continuum of progress does not mean it lacks predictive power; it is just that the predictions come in the far more useful form of numbers that can be read off a dial or photograph.

13: [9] Feynman was not the first to propose that positive-energy states go forwards and negative-energy states go backwards in time; he was largely pre-figured by Stückelberg:

E. C. G. Stückelberg, *La Mécanique du point matériel en théorie de relativité et en théorie des quanta*, Helv. Phys. Acta **15** (1942) 23.

18: [10] R. H. Good, *Properties of the Dirac Matrices*, Rev. Mod. Phys. **27** (1955) 187–211.

23: [11] The standard repository of all knowledge about propagators and 'relativistic' quantum mechanics is the first volume of Bjorken and Drell's similarly-named pair of textbooks.

J. D. Bjorken, S. D. Drell, *Relativistic Quantum Mechanics*, McGraw–Hill 1964.

30: [12] B. Podolsky, K. S. Kunz, *Fundamentals of Electrodynamics*, Marcel Dekker, 1969; pp 106–7.

P. A. M. Dirac, *Discussion of the infinite distribution of electrons in the theory of positrons*, Proc. Camb. Philos. Soc. **30** (1934) 150–163.

2: [1] Four papers make up Feynman's prize-winning Q.E.D. theory:

R. P. Feynman, *The Theory Of Positrons*, Phys. Rev. **76** (1949) 749–759.

R. P. Feynman, *Space-Time Approach To Quantum Electrodynamics*, Phys. Rev. **76** (1949) 769–789.

R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (1950) 440–457.

R. P. Feynman, *An Operator Calculus Having Applications In Quantum Electrodynamics*, Phys. Rev. **84** (1951) 108–128.

2: [2] S. S. Schweber, *Feynman and the visualization of space-time processes*, Rev. Mod. Phys. **58** (1986) 449–508.

2: [3] L. Gårding, *The mathematics of invariant wave equations*, pp 102–142 in *Invariant Wave Equations (Proceedings of a conference at Erice, 1977)*, Springer–Verlag Lecture Notes in Physics no. 73, 1978.

2: [4] D. Zwanziger, *Method of Characteristics in the External Field Problem*, or, *How to recognise an acausal equation when you see one*, pp 143–145 ibid.

2: [5] L. Hörmander, *Linear Partial Differential Operators*, Springer–Verlag, 1963.

2: [6] See Chapter 5.

3: [7] An excellent account of the theory of Green's functions and generalised functions can be found in

D. H. Griffel, *Applied Functional Analysis*, Ellis Horwood, 1981.

13: [8] Karl Popper has rigorously shown that no system can predict its own activity (*The Open Universe; an argument for indeterminism*, Hutchinson, 1982; pp 68–77) using a variation on the Stopping Problem for Turing machines. The idea of predicting the activity or evolution of a quantum system step–by–step as it happens is nearly as absurd. There is no value to be found in retaining this relic of a discredited interpretation of problem solving in classical mechanics. On the other hand, just because a theory does not purport to predict a continuum of progress does not mean it lacks predictive power; it is just that the predictions come in the far more useful form of numbers that can be read off a dial or photograph.

13: [9] Feynman was not the first to propose that positive-energy states go forwards and negative-energy states go backwards in time; he was largely pre-figured by Stückelberg:

E. C. G. Stückelberg, *La Mécanique du point matériel en théorie de relativité et en théorie des quanta*, Helv. Phys. Acta **15** (1942) 23.

18: [10] R. H. Good, *Properties of the Dirac Matrices*, Rev. Mod. Phys. **27** (1955) 187–211.

23: [11] The standard repository of all knowledge about propagators and 'relativistic' quantum mechanics is the first volume of Bjorken and Drell's similarly-named pair of textbooks.

J. D. Bjorken, S. D. Drell, *Relativistic Quantum Mechanics*, McGraw–Hill 1964.

30: [12] B. Podolsky, K. S. Kunz, *Fundamentals of Electrodynamics*, Marcel Dekker, 1969; pp 106–7.

# Chapter 9

## A Conclusion

*and some doubts*

*'...it is imperative in science to doubt; it is absolutely necessary, for progress in science, to have uncertainty as a fundamental part of your nature. To make progress in understanding we must remain modest and allow that we do not know. Nothing is certain or proved beyond all doubt.'*

Richard Phillips Feynman.

The quantum mechanics of Schrödinger, Heisenberg, *et. al.* is unsatisfactory for at least two reasons: it is a theory with the wrong invariance group; and the only way to discuss space-time developments is in terms of statistical measures of location, that is, by approximation. Von Neumann's rigorous mathematical formulation of quantum mechanics does not survive any attempt to rectify these faults. Thus, one way of viewing this thesis is as a progressively more radical movement away from von Neumann's axioms – only in chapter 2 are they used unmodified.

As regards the pursuit of a theory with the 'right' symmetries and a clearer use of locality, chapter 3 is somewhat of an exception. For while the thought experiments of de Broglie and Einstein–Podolsky–Rosen are discussed, the results apply, if anywhere, to the way 'quantum' systems become 'classical' ones – the disappearance of interference fringes and distant correlations. The results are qualitative but do suggest the way in which this transformation may occur.

In chapter 4 I have tried to cover the main lines of research into Poincaré–relativistic quantum mechanics (with the exception of Q.E.D.), and hence locality. I have avoided field theories throughout for the unconvincing reason that I know too little about them – though what I have understood does not suggest that there is a solution of the problem of locality.

A reasonable set of conclusions from my analysis of previous proposals is:

i) There is no position observable.

ii) The use of infinite hypersurfaces is more a mathematical artifice than a reflection of physical reality.

iii) Despite the momentum representation Hilbert space found by Wigner, the consequence of the foregoing is that there will be no coordinate representation Hilbert space.

Of course, even if these are reasonable there is no requirement to make them the foundation of a research programme – there are other theoretical requirements that can be altered or omitted. Indeed, this is what is done in the works analysed in chapters 6 and 7. In chapter 6 I have formally presented a quantum mechanics in which the notion of complete precision even in a single theoretical calculation is abandoned: that is, the theory does not produce unambiguous numbers. The result is almost certainly not a scientific theory as it cannot be tested – if it is not, in fact, just the usual quantum mechanics in disguise.

In chapter 7 some recent work by Graham Derrick (based on an idea Dirac published in 1949) is considered. Here the hypersurface is a backward–in–time light-cone, so that evolution is generated by the momentum 4-vector. However, just as basing classical mechanics on light-cones, rather than constant-time hyperplanes, still gives the same outcome, there is no reason to suppose that using a different hypersurface in quantum mechanics will give a different theory. There remain a number of quite major difficulties in formulating quantum mechanics on backward light-cones; so I would not claim to have any really firm judgement to offer on this approach. My reason for not pursuing this proposal is, simply, that I could see no way of sensibly overcoming the numerous difficulties that beset it.

By giving up the idea of a position observable, any positional relevance of quantum mechanics derives from the wave-function as a function of space and time. The mathematics presented in chapter 5 (an elementary proof is offered, which I developed because the 1974 paper by Gerhard Hegerfeldt is less than clear) strongly suggests that the symmetry group should not merely be the orthochronous Poincaré group, but the full Poincaré group – including time reflections. This implies using the, so-called, negative energy solutions of the wave equation: but then, these already have a legitimate place in the Feynman–Stückelberg interpretation used in Q.E.D. The final piece of motivation for the development offered in chapter 8 is the desire to represent the process of a particle entering or leaving the experimental arena – something that occupies a volume of space and time, and so can not be represented as an initial-value problem.

Finite space-time volumes are as convincingly covariant as any of the species of hypersurface, but have the advantage of finite size and a straightforward interpretation (as electron guns, photographic plates, etc.). The mathematics by which these come into use is not remarkably different from that used by Feynman in deriving Q.E.D.: the Green's function is found for the Dirac equation (for spin-half particles), but it is now used to solve the inhomogeneous problem. In deriving the Green's function a different rule is used to that developed by Feynman: now both positive and negative 'energies' propagate forwards and backwards in time to accomplish the 'evolution' of both the electron and the positron. The solution is therefore determined both from the past to the future and *vice versa*. The amplitudes at the sources and sinks thereafter seem to have a reasonable interpretation.

One consequence of this approach is that the momentum of a particle is once again an independent quantity. Another is that a form of trajectory can be ascribed to quantum particles.

This is not a well-elaborated theory; all I have managed is to stagger a few paces in the direction of a Poincaré–relativistic quantum mechanics. There is, however, plenty of scope for the theory to be tested; especially by comparison with Q.E.D., with which there seems to be considerable conflict. If the exchange photon for electron-magnetic interaction is propagated by the advanced/retarded Green's function given in §4, it follows, seemingly, that no mass renormalisation is needed.

My ignorance of Q.E.D. is such that there may be some simple fact that sinks the whole edifice that I have built. A ready illustration of my ignorance is revealed in the title of §4 – I am told there is no wave equation for the photon, yet the wave equation leads (with Feynman's rule) to the propagator for exchange photons, which, multiplied by $\gamma^\nu A_\nu$ , is also used in Q.E.D. formulae involving photon emission and absorption.

The obvious next step is to venture a new version of Q.E.D. to see where it differs from the standard theory.

# Appendix A

## The concepts of angular momentum and the centre of gravity in relativistic mechanics

*A. Papapetrou*
*Praktika Akademias Athenon* **14** *(1939) 540–547*

*'The bra lay there, dead.'*

**The Ganymede Takeover** by Philip K. Dick

Translated by Helen Ferguson.

*Note that all errors and infelicities in the following must remain the fault of the present author.*

1. For an isolated material system the Centre of Gravity Principle says that the centre of gravity of the system moves uniformly in a straight line; and the Angular Momentum Principle is that angular momentum remains constant in relation to any point. With the same formulation both principles remain applicable in relativistic mechanics[1].

In Newtonian mechanics the centre of gravity and its trajectory are absolutely definite, without any dependence on the coordinate system in which the movement of the material system is described. The question now arises as to whether this independence also exists in relativistic mechanics. The answer to this will occupy the major part of this work.

First of all the centre of gravity and the angular momentum principles are summarized in a single conservation law. From this it is then deduced that, in general, the centre of gravity changes when the material system under observation has an internal angular momentum (that is, in relation to its centre of gravity); from this the behaviour of internal angular momentum under coordinate transformation can also be obtained.

2. The proof works in a similar way to that of the energy–momentum conservation law; therefore we intend first of all to remind ourselves briefly of how this latter law is shown. Working from the equations which the material tensor, $T_\alpha^\beta$, satisfies in special relativity,

$$\frac{\partial T_\alpha^\beta}{\partial x^\beta} = 0 \quad , \tag{1}$$

these are integrated on a hyperplane $x^4 = ict = \text{constant}$. If it is assumed that the material system has finite spatial size, the terms corresponding to the three spatial coordinates, $x^\beta$ $(\beta = 1,2,3)$ disappear from (1), leaving finally:

$$\begin{aligned} &\frac{\partial}{\partial t}\int T_\alpha^4\, dv = 0 \quad , \\ &dv = dx_1\, dx_2\, dx_3 \quad . \end{aligned} \tag{2}$$

The proof is completed using the equation, which can be proved by Gauss' Theorem, that the integrals in (2) behave like the covariant components of a four-vector under coordinate transformations

$$\int T_\alpha^4\, dv = icG_\alpha \quad . \tag{3}$$

Here the first three components of $G_\alpha$ give the momentum, while the fourth gives the energy:

$$\begin{aligned} (G_1, G_2, G_3) &= \overline{G} \quad , \\ G_4 &= \frac{iE}{c} \quad . \end{aligned} \tag{3a}$$

3. A third order tensor can now be introduced which satisfies equations of the form of (1). A constant reference point is chosen, let it be the point $(\xi_\alpha)$, and each world-point, $(x_\alpha)$, is then ordered by the vector:

$$l_\alpha = x_\alpha - \xi_\alpha \quad . \tag{4}$$

If the tensor

$$F_{\alpha\beta}^\gamma = l_\alpha T_\beta^\gamma - l_\beta T_\alpha^\gamma \tag{5}$$

is formed from $l_\alpha$ and $T_\alpha^\beta$, then (1) and (4) imply:

$$\frac{\partial F_{\alpha\beta}^\gamma}{\partial x^\gamma} = 0 \quad . \tag{6}$$

It follows from (6), exactly as in the energy–momentum theorem, that for $t = \text{constant}$. quantities derived by

$$J_{\alpha\beta} = -\frac{i}{c}\int \left(l_\alpha T_\beta^4 - l_\beta T_\alpha^4\right)\, dv \tag{7}$$

are independent of time:

$$\frac{\partial J_{\alpha\beta}}{\partial t} = 0 \quad . \tag{8}$$

Therefore, in a fixed coordinate system they are dependent only on the point of reference, $(\xi_\alpha)$. Further, it can be proved in a similar fashion to the energy–momentum law that the quantities $J_{\alpha\beta}$ behave like the components of a second order covariant tensor under coordinate transformations. This last characteristic has already been taken into account in (7) by the use of the usual tensor notation. The factor $-i/c$ was introduced in (7) to facilate the results that follow.

4. We pass on to the meaning of the tensor $J_{\alpha\beta}$. According to (7) it is derived from an antisymmetric tensor, with three purely spatial and three mixed (spatio-temporal) components. The mixed components, $T^4_\alpha$, of the material tensor, which occur in the spatial components of (7), are related to the momentum density by:

$$T^4_\alpha = icg_\alpha \qquad (\alpha = 1,2,3) \quad .$$

This implies in the case of $J_{23}$, for example:

$$J_{23} = \int (l_2 g_3 - l_3 g_2)\, dv = \int (\bar{l} \times \bar{g})_x \, dv \quad .$$

The three spatial components of $J_{\alpha\beta}$ are thus identical with the components of the usual 3-dimensional angular momentum vector of the material system (for the chosen reference point $\xi_\alpha$):

$$(J_{23}, J_{31}, J_{12}) = \int \bar{l} \times \bar{g}\, dv = \overline{J} \quad . \tag{9}$$

To interpret the mixed components, put

$$\xi_4 = ic\tau \quad ;$$

also introduce the mass density, $\rho$, and the total mass, $\mu$, of the system:

$$\begin{aligned} \rho &= -\frac{T^4_4}{c^2} \quad , \\ \mu &= \int \rho\, dv = \frac{E}{c^2} \quad . \end{aligned} \tag{10}$$

This implies that

$$J_{\alpha 4} = ic\left[\int \rho x_\alpha \, dv - \xi_\alpha \mu - (t-\tau)G_\alpha\right] \quad . \tag{11}$$

The integral in the bracket is related to the coordinates, $s_\alpha$ of the centre of gravity, which is defined by

$$\int \rho x_\alpha \, dv = s_\alpha \mu \qquad (s_\alpha = s_\alpha(t)) \quad . \tag{12}$$

This gives, finally

$$J_{\alpha 4} = ic\left[(s_\alpha - \xi_\alpha)\mu - (t-\tau)G_\alpha\right] \qquad (\alpha = 1,2,3) \quad , \tag{13}$$

or, more symmetrically,

$$\begin{aligned} J_{\alpha 4} &= (s_\alpha - \xi_\alpha)G_4 - (s_4 - \xi_4)G_\alpha \quad , \\ &(G_4 = ic\mu \quad \text{and} \quad s_4 = ict) \quad . \end{aligned} \tag{13a}$$

If it is assumed that $t = \tau$, it follows, for example from (11), that

$$J_{\alpha 4} = ic \int \rho(x_\alpha - \xi_\alpha)\, dv = icM_\alpha \quad . \tag{14}$$

The mixed components $J_{\alpha 4}$ are therefore, except for a factor of $ic$, identical with the static moments $M_\alpha$ of the material system for the reference point $(\xi_\alpha)$ and the instant of time $t = \tau$. In accordance with the results (9) and (14), $J_{\alpha\beta}$ will be described as the *moment tensor* of the material system.

Firstly, if the conservation law (8) is applied to (9), it follows that the angular momentum of the material system remains constant:

$$\overline{J} = \int \bar{l} \times \bar{g}\, dv = \text{ constant} \quad . \tag{15}$$

It then follows from (13), as on the right-hand side only the quantity $s_\alpha$ depends on $t$:

$$\mu \frac{ds_\alpha}{dt} = G_\alpha \qquad (\alpha = 1,2,3) \quad , \tag{16}$$

these relations then give the Centre of Gravity Principle. From the foregoing it has been shown that the Centre of Gravity Principle and the Angular Momentum Principle are simultaneously contained in the conservation law (8).

5. Comparing (16) with

$$\mu \frac{ds_4}{dt} = ic\mu = G_4$$

shows that the centre of gravity moves on one of the straight lines parallel to the 4-vector $G_\alpha$, which can be described as the centre of gravity line of the material system. Let a point on the centre of gravity line be chosen as a reference point:

$$\xi_\alpha = s^*_\alpha \quad .$$

Then for $t = t^*$:

$$s_\alpha = s^*_\alpha = \xi_\alpha \quad ;$$

consequently, according to (13a):

$$J_{\alpha 4} = 0 \quad , \tag{17}$$

which, because of (8), also applies for all $t$. Conversely, if the equations (17) are fulfilled for a reference point then this lies on the centre of gravity line. If $t = \tau$ is taken then first of all this gives

$$ic\tau = ict \quad \text{or} \quad s_4 = \xi_4 \quad ,$$

while, on the other hand, because (17) is assumed, then according to (13):

$$s_\alpha = \xi_\alpha \qquad (\alpha = 1,2,3) \quad ;$$

thus $(\xi_\alpha)$ is the position of the centre of gravity at a certain point in time $t = \tau$. This brings us to the statement of the principle: the centre of gravity line is the geometrical location of the world-point in relation to which the mixed components of the momentum tensor disappear. The conditions (17) are thus equivalent to the equations of the centre of gravity line.

We will continue to consider how the components $J_{\alpha\beta}$ change with the reference point, $(\xi_\alpha)$. Let $(\xi_\alpha + \delta\xi_\alpha)$ be the new reference point. It follows immediately from (7) that

$$J_{\alpha\beta}(\xi + \delta\xi) = J_{\alpha\beta}(\xi) + G_\alpha \delta\xi_\beta - G_\beta \delta\xi_\alpha \quad . \tag{18}$$

It follows from this that the increment $\delta J_{\alpha\beta}$ disappears for a shift $(\delta\xi_\alpha)$ parallel to the vector $(G_\alpha)$: on any straight line parallel to the vector $(G_\alpha)$ the moment tensor remains unchanged. From the result just established, the characteristic of the centre of gravity line, previously found and expressed in (17), is supplemented in the following way: the angular momentum remains unchanged on the centre of gravity line.

Let $x_o$ be the coordinate system in which the momentum of the material system vanishes[1]:

$$\begin{aligned} G_{o1} = G_{o2} = G_{o3} &= 0 \quad , \\ G_{o4} = i\frac{E_o}{c} &= i\mu c \quad . \end{aligned} \tag{19}$$

[1] The existence of $x_o$ is equivalent to the demand for a time-like $(G_\alpha)$:

$$G_x^2 + G_y^2 + G_z^2 - \frac{E^2}{c^2} < 0 \quad , \tag{$\alpha$}$$

and this is the condition for the energy in any coordinate system to remain constant. The opposite to $(\alpha)$ also appears consistent with (1), but probably has no physical meaning.

In $x_o$ the material system is at rest as a whole and experiences only an internal motion. It follows immediately that in relation to this motion the angular momentum of the system is independent of the reference point:

$$\overline{J}(\xi + \delta\xi) = \overline{J}(\xi) = \overline{J}_o \quad . \tag{20}$$

This purely internal angular momentum is significant in the following observations.

6. We come finally to an examination of the question of how the centre of gravity and the angular momentum behave under coordinate transformations. To do this, it is obviously sufficient to examine the transition from a system at rest, $x_o$, to a system, $x$, in relation to which the rest frame, $x_o$, moves with relative velocity $\bar{v} = (v, 0, 0)$. For the sake of simplicity we will accept that the origin of coordinates in $x_o$ coincides with the centre of gravity. Then the centre of gravity line in $x_o$ is identical with the $x_{o4}$-axis, so that for some point on this axis the moment tensor will, because of (17), take the form:

$$J_{23} = J_{ox} \quad J_{31} = J_{oy} \quad J_{12} = J_{oz} \quad \text{otherwise} \quad J_{\alpha\beta} = 0 \quad . \tag{21}$$

We will calculate the moment tensor for the same reference point in coordinate system $x$. The transformation formula is:

$$\begin{aligned} x_1 &= \frac{x_{o1} - i\beta x_{o4}}{\sqrt{1-\beta^2}} \\ x_2 &= x_{o2} \qquad x_3 = x_{o3} \\ x_4 &= \frac{i\beta x_{o1} + x_{o4}}{\sqrt{1-\beta^2}} \end{aligned} \tag{22}$$

From this it follows that the quantities (21) transform like the products of the appropriate coordinates:

$$\begin{aligned} J_{23} &= J_{ox} \\ J_{31} &= \frac{J_{oy}}{\sqrt{1-\beta^2}} \\ J_{12} &= \frac{J_{oz}}{\sqrt{1-\beta^2}} \\ J_{14} &= 0 \\ J_{24} &= -\frac{i\beta J_{oz}}{\sqrt{1-\beta^2}} \\ J_{34} &= +\frac{i\beta J_{oy}}{\sqrt{1-\beta^2}} \end{aligned} \tag{23}$$

The components $J_{\alpha 4}$ are now in general different from zero, thus the centre of gravity line in $x_o$ is different from the centre of gravity line in $x$: *there is no definite world line which could describe the motion of the centre of gravity.*

The centre of gravity line can easily be characterised for the coordinate system $x$. A displacement, $(\delta\xi_\alpha)$ such as in (18), which leads to the disappearance of $J_{\alpha 4}$ is all that is needed for this. If it is borne in mind that, according to (19) and (22),

$$\begin{aligned} G_1 &= \frac{\mu_o v}{\sqrt{1-\beta^2}} \\ G_2 &= G_3 = 0 \\ G_4 &= \frac{i\mu_o c}{\sqrt{1-\beta^2}} \end{aligned} \tag{24}$$

(18) and (23) imply

$$\begin{aligned} J_{14}(\delta\xi) &= -\delta\xi_1 G_4 + G_1 \delta\xi_4 \\ J_{24}(\delta\xi) &= -\frac{i\beta J_{oz}}{\sqrt{1-\beta^2}} - \delta\xi_2 G_4 \\ J_{34}(\delta\xi) &= +\frac{i\beta J_{oy}}{\sqrt{1-\beta^2}} - \delta\xi_3 G_4 \end{aligned}$$

The conditions $J_{\alpha 4} = 0$ can thus be fulfilled by the following displacement:

$$\begin{aligned} \delta\xi_2 = s_y &= -\frac{v J_{ox}}{\mu_o c^2} \\ \delta\xi_3 = s_z &= +\frac{v J_{oy}}{\mu_o c^2} \\ \delta\xi_1 = \delta\xi_4 &= 0 \end{aligned} \tag{25}$$

The centre of gravity line in $x$ goes through the point (25) and is parallel to the vector $(G_\alpha)$. However, the displacement (25) is orthogonal to $(G_\alpha)$, so that (25) produces directly the actual displacement of the centre of gravity line in the transition $x_o \rightarrow x$. (25) can also be written as a vector:

$$\bar{s} = \frac{\bar{v} \times \overline{J}_o}{E_o} \quad . \tag{25a}$$

Next we will calculate the change in the angular momentum when the displacement (25) occurs. In other words, the internal angular momentum in coordinate system $x$. From (18) and (23)

$$\begin{aligned} J_{23}(\delta\xi) &= J_{ox} \\ J_{31}(\delta\xi) &= \frac{J_{oy}}{\sqrt{1-\beta^2}} - s_z G_1 = J_{oy}\sqrt{1-\beta^2} \\ J_{12}(\delta\xi) &= J_{oz}\sqrt{1-\beta^2} \end{aligned} \tag{26}$$

results. These formulae can be written more clearly if the internal angular momentum is decomposed into components orthogonal and parallel to $\bar{v}$:

$$\begin{aligned} J_{||} &= J_{o||} \\ J_{\perp} &= J_{o\perp}\sqrt{1-\beta^2} \end{aligned} \tag{26a}$$

7. Formula (25a) shows that with a given $\bar{v}$ the largest displacement of the centre of gravity line corresponds to the case $\bar{v} \perp \overline{J}_o$, where the following also applies:

$$s = \frac{v J_o}{\mu_o c^2} \qquad \bar{s} \perp \overline{J}_o \quad .$$

When it is taken into account[1] that $v \leq c$ the following result may be deduced: all possible lines of the centre of gravity of a given material system form a cylinder whose axis is the centre of gravity line of the purely internal motion (that is the $x_o$-axis) while the bounding surface lies orthogonal to $\overline{J}_o$ in three-dimensional space, and has the following radius:

$$r = \frac{J_o}{\mu_o c} \quad . \tag{27}$$

If the material system has no internal angular momentum it follows from (27), or else directly from (25a), that its centre of gravity is independent of the coordinate system: its motion is described by a definite world-line. In any other case there is a range characterised by the size of (27) for the position of the centre of gravity.

---

References with the page number on which they occur

2: [1] Cf. M. von Laue *The theory of relativity* **1**, §227, Vieweg, 1921. The Centre of Gravity Principle can be proved in a similar way.

[1] If the speeds within the system are of the order of $v$, and the linear dimensions of the system are of the order of $R$, the following applies:

$$J_o \approx \mu_o v R \qquad s \approx \frac{v^2}{c^2} R = \beta^2 R \quad .$$

If $\beta \ll 1$, the displacements of the centre of gravity tends, like $\beta^2$, to zero; and the same applies, from (26), to the change in internal angular momentum. This statement provides the connection with Newtonian mechanics, where the centre of gravity and the angular momentum are independent of the coordinate system.

# Appendix B

## The construction of definite expressions for the particle density of the Klein–Gordon field

*Bernd Gerlach, Dieter Gromes, Joachim Petzold*
*Zeitschrift für Physik* **204** *(1967) 1–11*

*'In the midst of the word he was trying to say*
*In the midst of his laughter and glee,*
*He had softly and suddenly vanished away —*
*For the Snark was a Boojum, you see.'*

Translated by Helen Ferguson.

*Note that all errors and infelicities in the following must remain the fault of the present author.*

## §1 Introduction

In a previous work[1] it was shown that the 4–vector assigned to the free Klein–Gordon field, $\psi(x)$,*

$$s^{\mu}(x) = ie\left(\overline{\psi(x)}\psi^{|\mu}(x) - \psi(x)\overline{\psi^{|\mu}}(x)\right) \tag{1.1}$$

cannot be interpreted as an electric current density, as the 0–component of this vector itself is not positive-definite, even if for $\psi(x)$ only solutions of the Klein–Gordon equation for positive frequencies are admitted. This poses the problem of finding a more suitable definition of the current density. (Infinitely many) quantities, $\tilde{s}^{\mu}(x)$, will be given which are:

(1) are real,
(2) transform like 4–vectors,
(3) are expressed bilinearly through the Klein–Gordon field,
(4) fulfil the continuity equation

$$s^{\mu}_{|\mu}(x) = 0 \quad ,$$

(5) and, moreover, have a positive-definite 0–component if limited to positive frequency solutions.

Given these requirements, it can be seen that $\tilde{s}^{\mu}$ goes over to the current density of the Schrödinger equation in the limiting 'non-relativistic' case. The connection between current and field is now completely non-local.

## §2 The general form of the particle density

The solutions, $\psi^{+}(x)$, of the free Klein–Gordon equation for positive frequencies can be generally written in the form

$$\psi^{+}(x) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int a(\mathbf{k})e^{-ik_{\nu}x^{\nu}}\frac{d^3k}{2k_0} \qquad \text{with} \quad k_{\rho}k^{\rho} = \kappa^2 \quad . \tag{2.1}$$

The translations $\psi(x) \rightarrow \psi(x+y)$ can then be generated through the transformation

$$a(\mathbf{k}) \rightarrow a(\mathbf{k})e^{-ik_{\nu}y^{\nu}} \quad . \tag{2.2}$$

In the context of the solutions (2.1) the usual expression can be seen as a bilinear function of $a(\mathbf{k})$:

$$s^{\mu}_{+}(x) = \frac{q_e}{(2\pi)^3}\int (k^{\mu}+k'^{\mu})\overline{a(\mathbf{k})}a(\mathbf{k}')e^{-i(k_{\nu}-k'_{\nu})x^{\nu}}\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0} \quad . \tag{2.3}$$

If it is required that the total charge should be equal to the elementary charge, $q_e$,

$$\int s^{0}_{+}(x)\,d^3x = q_e\int |a(\mathbf{k})|^2\,\frac{d^3k}{2k_0} \quad ,$$

* We are using the metric $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$. This further means that

$$\kappa = \frac{mc}{h} \qquad k_0 = \sqrt{\kappa^2+\mathbf{k}^2} \qquad \psi_{|\mu} = \frac{\partial\psi}{\partial x^{\mu}} \quad .$$

the normalization condition

$$\int |a(\mathbf{k})|^2 \frac{d^3k}{2k_0} = 1 \quad , \tag{2.4}$$

follows for $a(\mathbf{k})$. To establish first of all the form of current density, $\tilde{s}^\mu_+$, which we are concerned with, we will work from requirement 3 (§1), according to which it should be a bilinear function of the Klein–Gordon field (2.1). Considering the field as expressed by $a(\mathbf{k})$, the most general bilinear formulation is established by analogy to (2.3):

$$\begin{aligned}\tilde{s}^\mu_+(x) = &\int F^\mu(\mathbf{k},\mathbf{k}',x)\overline{a(\mathbf{k})}a(\mathbf{k}')\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0} \\ &+ \int G^\mu(\mathbf{k},\mathbf{k}',x)a(\mathbf{k})a(\mathbf{k}')\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0} \\ &+ \int H^\mu(\mathbf{k},\mathbf{k}',x)\overline{a(\mathbf{k})a(\mathbf{k}')}\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0}\end{aligned} \tag{2.5}$$

and an attempt is made to fulfil the other characteristics which are demanded of $\tilde{s}^\mu_+$ in §1, through a suitable choice of $F^\mu$, $G^\mu$, $H^\mu$. The functions $F$, $G$, $H$ should be continuous. For the purposes of discussion, a $\delta$–distribution can also be introduced into (2.5) in place of $a(\mathbf{k})$. Because of the transformation characteristic (2.2) of $a(\mathbf{k})$, the dependence of $F$, $G$, $H$ on $x$ can be written down straight-away:

$$\begin{aligned}F^\mu(\mathbf{k},\mathbf{k}',x) &= F^\mu(\mathbf{k},\mathbf{k}')e^{i(k_\nu - k'_\nu)x^\nu} \quad , \\ G^\mu(\mathbf{k},\mathbf{k}',x) &= G^\mu(\mathbf{k},\mathbf{k}')e^{-i(k_\nu + k'_\nu)x^\nu} \quad , \\ H^\mu(\mathbf{k},\mathbf{k}',x) &= H^\mu(\mathbf{k},\mathbf{k}')e^{i(k_\nu + k'_\nu)x^\nu} \quad .\end{aligned}$$

The scalar quantity $k_\rho k'^\rho$ and the two independent 4–vectors $k^\mu + k'^\mu$ and $k^\mu - k'^\mu$ are formed from $\mathbf{k}$ and $\mathbf{k}'$ (the scalars $k^\rho k_\rho = \kappa^2$ and $k'^\rho k'_\rho = \kappa^2$ are independent of $\mathbf{k}$ and $\mathbf{k}'$). Bearing in mind (2.5), only the part of $G$ and $H$ symmetric with respect to $\mathbf{k}$ and $\mathbf{k}'$ need be taken into account, we can write

$$\begin{aligned}F^\mu(\mathbf{k},\mathbf{k}') &= (k^\mu + k'^\mu)F_1(k_\rho k^{\rho\prime}) + (k^\mu - k'^\mu)F_2(k_\rho k^{\rho\prime}) \quad , \\ G^\mu(\mathbf{k},\mathbf{k}') &= (k^\mu + k'^\mu)G_2(k_\rho k^{\rho\prime}) \quad , \\ H^\mu(\mathbf{k},\mathbf{k}') &= (k^\mu + k'^\mu)H_2(k_\rho k^{\rho\prime}) \quad .\end{aligned} \tag{2.6}$$

If $\tilde{s}^\mu_+(x)$ is split up by $\tilde{s}^\mu_+(x) = \tilde{s}^\mu_{(1)}(x) + \tilde{s}^\mu_{(2)}(x)$, where $\tilde{s}^\mu_{(1)}(x)$ already satisfies the continuity equation $\tilde{s}^\mu_{(1)|\mu}(x) = 0$, given which the requirement that $\tilde{s}^\mu_{(2)}$ be divergence free imposes conditions on the functions $F_2$, $G_2$, $H_2$. In particular the following must hold at the origin of coordinates:

$$\begin{aligned}0 = \frac{1}{i}\tilde{s}^\mu_{(2)|\mu}(0) = &\int (k-k')^2 F_2(k_\rho k^{\rho\prime})\overline{a(\mathbf{k})}a(\mathbf{k}')\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0} - \\ &\int (k+k')^2 G_2(k_\rho k^{\rho\prime})a(\mathbf{k})a(\mathbf{k}')\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0} + \\ &\int (k+k')^2 H_2(k_\rho k^{\rho\prime})\overline{a(\mathbf{k})a(\mathbf{k}')}\frac{d^3k}{2k_0}\frac{d^3k'}{2k'_0}\end{aligned} \tag{2.7}$$

Where the convention $q_\nu q^\nu = (q)^2$ is used. (2.7) should be satisfied for all $a(\mathbf{k})$, so that taking the special case:

$$a(\mathbf{k}) = A2k^0_{(1)}\delta(\mathbf{k} - \mathbf{k}_{(1)}) + B2k^0_{(2)}\delta(\mathbf{k} - \mathbf{k}_{(2)}) \tag{2.8}$$

(to be more precise, the $\delta$–function should be replaced by a sequence of regular functions, $f_\nu$, with $f_\nu \to \delta$ converging in the distributional sense.). Then

$$\begin{aligned}&(\overline{A}B + A\overline{B})F_2(k_{(1)\rho}k^{\rho\prime}_{(2)})(k_{(1)} - k_{(2)})^2 - \\ &A^2 G_2(\kappa^2)4\kappa^2 - B^2 G_2(\kappa^2)4\kappa^2 - 2ABG_2(k_{(1)\rho}k^{\rho\prime}_{(2)})(k_{(1)} + k_{(2)})^2 + \\ &\overline{A}^2 H_2(\kappa^2)4\kappa^2 - \overline{B}^2 H_2(\kappa^2)4\kappa^2 + 2\overline{A}\overline{B}H_2(k_{(1)\rho}k^{\rho\prime}_{(2)})(k_{(1)} + k_{(2)})^2 = 0\end{aligned} \tag{2.9}$$

is obtained from (2.7). If the case $B = 0$ is considered, and $A^2$ is chosen first to be real and then pure imaginary, $G_2(\kappa^2) = 0 = H_2(\kappa^2)$ is obtained. If the case $A = B$ is considered, and $A$ is first chosen to be real, then pure imaginary, and then such that $A^2$ is pure imaginary, (2.9) gives:

$$\begin{aligned}F_2(k^\rho_{(1)}k'_{(2)\rho})(k_{(1)} - k_{(2)})^2 &= G_2(k^\rho_{(1)}k'_{(2)\rho})(k_{(1)} + k_{(2)})^2 \\ &= H_2(k^\rho_{(1)}k'_{(2)\rho})(k_{(1)} + k_{(2)})^2 = 0 \quad .\end{aligned} \tag{2.10}$$

Now for $\mathbf{k}_{(1)} \neq \mathbf{k}_{(2)}$, $(k_{(1)} + k_{(2)})^2 \geq 4\kappa^2$ and $(k_{(1)} - k_{(2)})^2 < 0$. Therefore from (2.10), $F_2 = G_2 = H_2 = 0$ (for $F_2(k^\rho k'_\rho)$ this follows immediately for $\mathbf{k} \neq \mathbf{k}'$, however because $F_2$ is continuous then it also follows for $\mathbf{k} = \mathbf{k}'$, that is, for $F_2(\kappa^2)$).

If (2.5) is finally tackled again, writing the remaining term using (2.6) as $F^\mu = (k^\mu + k'^\mu)F_1$, it can be seen that the function $F_1$ must be real in order for $\tilde{s}^\mu_+$ to be real. If now one writes

$$F_1 = \frac{q_e}{(2\pi)^3} f \quad ,$$

the new expression for the current takes the form

$$\tilde{s}^\mu_+(x) = \frac{q_e}{(2\pi)^3} \int (k^\mu + k'^\mu) f(k^\rho k'_\rho) \overline{a(\mathbf{k})} a(\mathbf{k}') e^{-i(k_\nu - k'_\nu)x^\nu} \frac{d^3k}{2k_0} \frac{d^3k'}{2k'_0} \tag{2.11}$$

with a real $f$. Because $k^\rho k'_\rho \geq \kappa^2$, the argument of $f$ always lies in the interval $[\kappa^2, \infty)$. The total charge becomes

$$\int \tilde{s}^0_+(x)\, d^3x = q_e f(\kappa^2) \int |a(\mathbf{k})|^2 \frac{d^3k}{2k_0} \quad . \tag{2.12}$$

If the normalization (2.4) is retained for $a(\mathbf{k})$, the normalization rule $f(\kappa^2) = 1$ follows for $f$. The total charge is independent of the function $f$!

The limiting 'non-relativistic' case is characterized by $a(\mathbf{k})$ only being appreciably different from 0 for values of $\mathbf{k}$ with $|\mathbf{k}| \ll \kappa = mc/\hbar$. For $\mathbf{k}$ and $\mathbf{k}'$ of this type, $f(k^\rho k'_\rho) \approx f(\kappa^2) = 1$. Thus $\tilde{s}^\mu_+(x)$, given by (2.11), becomes $s^\mu_+(x)$, given by (2.3) (and this is recognizable as the relevant expression for the Schrödinger equation)*.

## §3 Explicit forms of definite particle densities

Given the general structure of $\tilde{s}^\mu_+(x)$, demonstrated in (2.5), and assuming the requirements 1–4 to be fulfilled, the next task is to state a function $f$ such that $\tilde{s}^0_+(x) \geq 0$ applies for all $a(\mathbf{k})$ and any $x^\nu$. As $\tilde{s}^0_+(x)$ must be greater than or equal to zero for all $a(\mathbf{k})$ which fulfil the normalization (2.4), this must also apply to the special case of $\hat{a}(\mathbf{k}) = a(\mathbf{k})e^{-ik_\nu x^\nu}$ for fixed $x^\nu$:

$$\hat{s}^0 = \int (k^0 + k'^0) f(k^\rho k'_\rho) \overline{\hat{a}(\mathbf{k})} \hat{a}(\mathbf{k}') \frac{d^3k}{2k_0} \frac{d^3k'}{2k'_0} \geq 0 \quad . \tag{3.1}$$

If, conversely, (3.1) is correct for all $\hat{a}(\mathbf{k})$, it also applies for the above special case, and thus it is always the case that $x\tilde{s}^0_+(x) \geq 0$.

So, if $\tilde{s}^0_+$ is positive definite at a single world point then the same is true at all other points.

It will now be shown that the inequality (3.1) is satisfied by the infinite set

$$f = f_\nu(k^\rho k'_\rho) = \left( \frac{2\kappa^2}{k^\rho k'_\rho + \kappa^2} \right)^{1+\nu} \tag{3.2}$$

where $\nu = 1, 2, 3, \ldots$ can be used†.

---

* Conversely, it follows from this that the 'relativistic' generalization is by no means unambiguous for 'non-relativistic' observables.

† For functions of this kind one is lead to investigate the case of a single spatial dimension, where the simplifying transformation $k = \kappa \sinh u$ can be carried out. This then gives

$$k^0 k'^0 - k^1 k'^1 = \kappa^2 \left( 2\cosh^2 \frac{u - u'}{2} - 1 \right) \quad ,$$

$$k^0 + k'^0 = 2\kappa \cosh \frac{u - u'}{2} \left( \cosh \frac{u}{2} \cosh \frac{u'}{2} + \sinh \frac{u}{2} \sinh \frac{u'}{2} \right) \quad ,$$

and if $\cosh u$ or $\sinh u$ is combined with the state function $\hat{a}$, the sum of two integral operations is produced whose kernel still depends only on $u - u'$, so that the Fourier transform approach can be used to solve it.

Appendix B : *Explicit forms of definite particle densities*

In (3.1) the functions $\hat{a}(\mathbf{k}) = \hat{a}(k,\theta,\phi)$ are decomposed using the (orthonormalized) spherical harmonics:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(l-|m|)!\,(2l+1)}{(l+|m|)!\,4\pi}}\,P_l^{|m|}(\cos\theta)e^{im\phi} \quad .$$

$$\hat{a}(\mathbf{k}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{+l} a_{lm}(k)Y_l^m(\theta,\phi) \quad .$$

Substituting this into equation (3.1), the following is produced:

$$\hat{s}^0 = \sum_{l,m,l',m'}\int (k^0+k'^0)g_{l,l',m,m'}(k,k')\overline{a_{l,m}(k)}a_{l',m'}(k')\frac{k^2\,dk}{2k_0}\frac{k'^2\,dk'}{2k'_0} \quad , \tag{3.3}$$

in which

$$g_{l,l',m,m'}(k,k') = \int f(k^0k'^0 - kk'\cos\beta)\overline{Y_l^m(\theta,\phi)}Y_{l'}^{m'}(\theta',\phi')\,d(\cos\theta)\,d\phi\,d(\cos\theta')\,d\phi' \tag{3.4}$$

is introduced with $\cos\beta = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi-\phi')$. Similarly, $f$ can be decomposed in terms of the spherical harmonic functions:

$$f(k^0k'^0 - kk'\cos\beta) = \sum_{l=0}^{\infty} f_l(k,k')Y_l^0(\beta) \quad ,$$

which, by the Addition theorem for spherical harmonics[2] produces

$$f(k^0k'^0 - kk'\cos\beta) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{\sqrt{4\pi}f_l(k,k')}{\sqrt{2l+1}}Y_l^m(\theta,\phi)\overline{Y_l^m(\theta',\phi')} \quad .$$

If this expression is substituted into (3.4), the following is produced:

$$\begin{aligned} g_{l,l',m,m'}(k,k') &= \sum_{l''=0}^{\infty}\sum_{m=-l''}^{l''}\frac{\sqrt{4\pi}f_l''(k,k')}{\sqrt{2l''+1}}\delta_{l',l''}\,\delta_{m',m''}\,\delta_{l,l''}\,\delta_{m,m''} \\ &= \frac{\sqrt{4\pi}f_l(k,k')}{\sqrt{2l+1}}\delta_{l,l'}\,\delta_{m,m'} \\ &= 2\pi\delta_{l,l'}\,\delta_{m,m'}\int_{-1}^{+1} f(k^0k'^0 - kk'\cos\beta)P_l(\cos\beta)\,d(\cos\beta) \quad . \end{aligned}$$

Accordingly in (3.3) only orthogonal terms relating to $l$ and $m$ remain, and as the $a_{lm}(k)$ can be chosen independently of each other, it follows as a necessary and sufficient condition that the kernels

$$K_l(k,k') = (k^0+k'^0)\int_{-1}^{+1} f(k^0k'^0 - kk't)P_l(t)\,dt \quad \text{for} \quad l = 0,1,2,\ldots \tag{3.5}$$

must be positive-definite.

If for $f$ we write $f_\nu$ as introduced in (3.2), then first of all (3.5) gives

$$K_l(k,k') = \frac{(2\kappa^2)^{1+\nu}}{(k)^{1+\nu}(k')^{1+\nu}}(k^0+k'^0)\int_{-1}^{+1}\frac{P_l(t)}{\left(\frac{\kappa^2+k^0k'^0}{kk'} - t\right)^{1+\nu}}\,dt \quad . \tag{3.6}$$

Now in the case that $z$ does not lie on the real axis between $-1$ and $+1$, which is so because

$$z = \frac{\kappa^2 + k^0k'^0}{kk'} > 1 \quad ,$$

the following formula applies[3].

$$Q_l(z) = \frac{1}{2}\int_{-1}^{+1}\frac{P_l(t)\,dt}{z-t}$$

Differentiating this $\nu$ times we obtain the relation[4]

$$\int_{-1}^{+1} \frac{P_l(t)\,dt}{(z-t)^{1+\nu}} = \frac{2(-1)^\nu}{\nu!}\frac{d^\nu}{dz^\nu}Q_l(z) = \frac{2(-1)^\nu}{\nu!\,(z^2-1)^{\frac{\nu}{2}}}Q_l^\nu(z)$$

producing the following representation for the kernel, $K_l$:

$$K_l(k,k') = \frac{2(2\kappa^2)^{1+\nu}(-1)^\nu(k^0+k'^0)}{\nu!\,(k)^{1+\nu}(k')^{1+\nu}\left[\left(\frac{\kappa^2+k^0k'^0}{kk'}\right)^2-1\right]^{\frac{\nu}{2}}}Q_l^\nu\left(\frac{\kappa^2+k^0k'^0}{kk'}\right) \quad .$$

It thus remains to investigate whether the inequality

$$\int_0^\infty\int_0^\infty \frac{(-1)^\nu(k^0+k'^0)}{(k)^{1+\nu}(k')^{1+\nu}\left[\left(\frac{\kappa^2+k^0k'^0}{kk'}\right)^2-1\right]^{\frac{\nu}{2}}}Q_l^\nu\left(\frac{\kappa^2+k^0k'^0}{kk'}\right)\overline{a(k)}a(k')\,dk\,dk' \geq 0$$

is satisfied. Making the substitution

$$k = \frac{\kappa}{\sinh u} \quad , \qquad k^0 = \kappa\coth u$$

this may be re-expressed as

$$(-1)^\nu\int_0^\infty\int_0^\infty (\sinh(u+u'))^{1-\nu}\,Q_l^\nu(\cosh(u+u'))\overline{b(u)}b(u')\,du\,du' \geq 0 \quad . \tag{3.7}$$

Where $b(u) = \kappa^{\frac{1}{2}-\nu}a(k)\cosh u(\sinh u)^{\nu-2}$. The kernel henceforth depends only on the sum of the argument, so that the method of Laplace transforms, that is the expansion of exponential functions as series, suggests itself as the next step.

In order to do this, consider the formula[5]

$$Q_l^\nu(x+i0) = (-1)^\nu e^{i\nu\pi/2}\left(Q_l^\nu(x) - \frac{i\pi}{2}P_l^\nu(x)\right) \quad , \qquad -1 < x < 1 \quad .$$

If trigonometric substitutions are made to give $P_l^\nu(\cos\theta)$ and $Q_l^\nu(\cos\theta)$[6], then the following is produced:

$$Q_l^\nu(\cos\theta+i0) = \sqrt{\pi}(-2)^\nu e^{i\nu\pi/2}(\sin\theta)^\nu\frac{\Gamma(l+\nu+1)}{\Gamma\left(l+\frac{3}{2}\right)}\sum_{n=0}^\infty c_n e^{-i(2n+l+\nu+1)\theta} \tag{3.8}$$

with

$$c_n = \frac{\left(\frac{1}{2}+\nu\right)_n(1+l+\nu)_n}{n!\,\left(l+\frac{3}{2}\right)_n} \quad ,$$

valid for $0 < \theta < \pi$.

In this context, define

$$(\alpha)_n = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} = \alpha(\alpha+1)\ldots(\alpha+n-1) \quad .$$

$Q_l^\nu(\cos\theta+i0)$ is single-valued and analytic in $\theta$ as long as $\cos\theta$ lies in the upper half-plane, or on the real axis except where $\cos\theta = \pm 1$. The coefficients $c_n$ behave for large $n$ like $n^{2\nu-1}$. Thus the analytic continuation of (3.8) gives

$$\begin{aligned} Q_l^\nu(\cosh y+i0) &= Q_l^\nu(\cosh y) \\ &= \sqrt{\pi}(-2)^\nu(\sinh y)^\nu\frac{\Gamma(l+\nu+1)}{\Gamma\left(l+\frac{3}{2}\right)}\sum_{n=0}^\infty c_n e^{-(2n+l+\nu+1)y} \quad \text{for} \quad y > 0 \quad . \end{aligned} \tag{3.9}$$

(3.7) imposes the following condition on (3.9)

$$(-1)^{\nu}(\sinh y)^{1-\nu}\,\mathcal{Q}_l^{\nu}(\cosh y) = \sqrt{\pi}2^{\nu}(\sinh y)^{\nu}\frac{\Gamma(l+\nu+1)}{\Gamma\left(l+\frac{3}{2}\right)}\sum_{n=0}^{\infty}c_n e^{-(2n+l+\nu+1)y}$$

$$= \sqrt{\pi}2^{\nu-1}\frac{\Gamma(l+\nu+1)}{\Gamma\left(l+\frac{3}{2}\right)}\sum_{n=0}^{\infty}(c_n - c_{n-1})e^{-(2n+l+\nu)y} \quad (c_{-1}=0). \tag{3.10}$$

It is now easy to calculate: $c_0 = 1$ and

$$c_n - c_{n-1} < 0 \quad \text{for} \quad \nu = 0\,, \quad n = 1,2,\ldots$$
$$c_n - c_{n-1} > 0 \quad \text{for} \quad \nu = 1,2,\ldots\,, \quad n = 1,2,\ldots$$

Thus for $\nu = 0$ the density becomes indefinite. On the other hand, for $\nu = 1,2,\ldots$ the expression (3.7) with (3.10) ($y = u + u'$ is assumed) can be written

$$\sum_{n=0}^{\infty}A_n^{(l,\nu)}\left|\int_0^{\infty}e^{-(2n+l+\nu)}b(u)\,du\right|^2$$

with $A_n^{(l,\nu)}$ positive, from which the definiteness is clear. Since, for a function $b(u)$ that is not equal to zero (basically, this means it does not disappear identically) not all integrals

$$\int_0^{\infty}e^{-(2n+l+\nu)}b(u)\,du$$

can disappear[7], and because all $A_n^{(l,\nu)}$ are greater than zero, $\tilde{s}_+^0$ is positive-definite for all world-points $x$ ($\tilde{s}_+^0(x) > 0$).

From the $f_\nu$ new functions that give more positive densities can be formed by the superposition using the formula

$$f(z) = \sum_{\nu=1}^{\infty}\alpha_\nu f_\nu(z) \quad \text{where} \quad \alpha_\nu \geq 0$$

In this way, functions of the form

$$f(k_\rho k'^\rho) = \left(\frac{(1+C)\kappa^2}{C\kappa^2 + k_\rho k'^\rho}\right)^{1+n} = (1+C)^{1+n}\sum_{\nu=n}^{\infty}\binom{\nu-1}{\nu-n}\frac{(1-C)^{\nu-n}}{2^{1+\nu}}f_\nu(k_\rho k'^\rho)$$

with $-1 < C \leq 1$ and $n = 1,2,\ldots$.

In relation to a work by Pais and Uhlenbeck[8], that has not yet been discussed, it is interesting to note that the function $f(k_\rho k'^\rho) = \exp(1 - k_\rho k'^\rho/\kappa^2)$ does not give a positive-definite density. This can be seen from the kernel formed from (3.5):

$$K_0(k,k') = 2(k^0 + k'^0)\exp\left(1 - \frac{k^0 k'^0}{\kappa^2}\right)\frac{\kappa^2}{kk'}\sinh\left(\frac{kk'}{\kappa^2}\right) \quad .$$

If

$$a(k) = \frac{A}{2}\delta(k) + \frac{B}{2}\delta(k - \tilde{k}) \quad ,$$

is taken, it follows that

$$\int K_0(k,k')\overline{a(k)}a(k')\,dk\,dk' = A^2\kappa + B^2\tilde{k}^0\exp\left(1 - \frac{(\tilde{k}^0)^2}{\kappa^2}\right)\frac{\kappa^2}{\tilde{k}^2} + AB(\kappa + \tilde{k}^0)\exp\left(1 - \frac{\tilde{k}^0}{\kappa}\right) \quad .$$

This quadratic form can be negative for small $\tilde{k}$.

The physical problems connected with the introduction of the suggested new current will be dealt with in a subsequent article. In particular it will be shown that the particle densities spread out causally.

*The appendix dealing with the 2–dimensional case has been omitted.*

---

References with the page number on which they occur

2: [1] B. Gerlach, D. Gromes, J. Petzold Z. Phys. **202** (1967) 401.
5: [2] A. Erdélyi, W. Magnus, F. Oberhettinger, F. G. Tricomi *Higher transcendental functions*, New York, McGraw–Hill 1953, vol. 1 §168(2).
5: [3] ibid. §154(29).
5: [4] ibid. §148(5).
6: [5] ibid. §144(9).
6: [6] ibid. §.146(2) and (3).
7: [7] G. Doetsch *Handbuch der Laplace–Transformation*, Birkhäuser, Basel 1950; vol. I, §74, theorem 4.
7: [8] A. Pais, G. E. Uhlenbeck Phys. Rev. **79** (1950) 145.