Pramana – J. Phys., Vol. 37, No. 3, September 1991, pp. 179–233.

Quantum theory in external electromagnetic and gravitational fields: A comparison of some conceptual issues

T PADMANABHAN

Theoretical Astrophysics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

MS received 3 July 1991

Abstract. The quantum theories of a scalar field interacting with external electromagnetic and gravitational fields respectively are compared. It is shown that several peculiar features, like the ambiguity of particle definition, thermal effects etc., which are thought to be special to quantum theory in curved spacetime, have analogues in the case of electromagnetism.

Keywords. Quantum theory; quantum gravity; Rindler frame; Hawking radiation; pair creation; expanding universe; back reaction.

PACS Nos 03·70; 04·60; 11·90

Table of contents

1. Introduction

In the study of two systems S_1 and S_2, which are interacting with each other, we can distinguish three limiting situations: The first one treats both the systems as classical and the classical equations of motion are used to describe them. In the second case both the systems are treated as quantum mechanical and the rules of quantum theory are used to describe them. For a class of systems, we may also have a third limit, viz. the one in which one of the systems say, S_1, is classical and S_2 is quantum mechanical. This limit is conceptually of a different kind compared to the other two, in the sense that we now have to couple a quantum system to a classical one. Since the language of quantum theory is very different from that of classical physics, this task is non-trivial.

There are, however, situations in which the third level of approximation is of importance. One such situation is when S_1 describes the gravitational field and S_2, some other matter field. The exact, quantum, description of such a system is not known. It is reasonable to hope that the study of the limit, in which a quantized field interacts with classical gravity, will provide us with some insight regarding the exact quantum theory. Because of this hope, considerable amount of work was done in investigating the behaviour of quantum field theory in curved spacetime (Birrel and Davies 1982). Though no useful insight regarding the nature of quantum gravity was gained, these investigations have uncovered several conceptual issues and surprises. Most of these aspects were believed to be rather special features, "somehow" related to the nature of gravity.

The purpose of this review is to look closely at some of these effects (which arise when classical gravity interacts with a quantum field) and compare them with corresponding situations in the case of a classical electromagnetic field interacting with a quantum field. We will see that there are several similarities between the two and that the results involving gravity are by no means special or mysterious.

Such a comparison also helps us in a different way. Since the exact theory of quantum electrodynamics is (believed to be) known, we should be able to resolve satisfactorily any conceptual issue which arises in the case of a classical electromagnetic field interacting with another quantum field. By using the analogy between the two fields (which we have previously established), we will be able to clarify the various conceptual difficulties encountered in the case of gravity. We shall also address this question in this review.

The review is structured as follows: Parts 2 and 3 summarize several pieces of background information (regarding path integrals and effective action) which are needed to study a classical system interacting with a quantum system. The discussion is limited to setting up the notation and highlighting the key results. The core of the review is contained in parts 4 and 5. Part 4 discusses the quantum theory of a charged scalar field in an external electromagnetic field; part 5 studies the corresponding situation of a quantized scalar field interacting with classical gravity. In both parts, we have chosen *specific* kinds of external fields in order to emphasize the analogy. We also provide a detailed discussion—in part 4— of pair creation in an electric field and renormalization of Euler–Heisenberg effective lagrangian. Part 6 summarizes the conclusions.

As should be clear from the above description, the review focusses on certain *specific* aspects of quantum theory in external fields. Several other interesting and related issues, like nature and validity of semiclassical approximation (Singh and Padmanabhan 1989; Padmanabhan 1989; Banks 1985; Hartle 1986), the issue of back reaction (Duff 1981; Ford 1982; Padmanabhan 1989), vacuum instability in different kinds of external fields (Ginzburg 1987), quantization of fermionic fields (Greiner *et al* 1985) etc are not discussed. The interested reader can find more on these topics in the references provided.

2. Path integral

In this part, we shall quickly summarize the key results from path integrals which are needed later; more detailed discussion of these topics are available in Feynman and Hibbs (1965), Shulman (1981) and Rivers (1987).

2.1 *Path integral techniques*

In classical mechanics, the laws governing the motion of a particle in a potential $V(x)$ can be obtained from the principle of least action. This principle states that the trajectory followed by a particle in travelling from (t_1, x_1) to (t_2, x_2) is the one which makes the action

$$A[x(t)] \equiv \int_{t_1}^{t_2} \mathrm{d}t L(\dot{x}, x) = \int_{t_1}^{t_2} \mathrm{d}t(\tfrac{1}{2}m\dot{x}^2 - V(x)) \tag{2.1}$$

an extremum. This prescription leads to the equation:

$$m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + V'(x) = 0 \tag{2.2}$$

which determines the extremum path. The solution to this differential equation connecting the events $\mathscr{P}1$ and $\mathscr{P}2$ gives the classical trajectory of the particle. We will denote this classical path by $x_c(t)$ and the corresponding value for the action, $A(x_c)$ by A_c.

The classical description of dynamics depends crucially on the existence of well defined trajectories for motion. To characterize a path at any instant of time, it is necessary to specify both the position and velocity of the particle at that instant of time. Since uncertainty principle forbids such a simultaneous specification of position and momentum the above description needs to be modified in quantum mechanics.

A suitable modification can be arrived at by considering the results of standard two-slit interference experiment with, say, electrons. These experiments suggest that the electrons do not follow a definite trajectory in travelling from the electron gun to the screen. Instead, we must associate with each path connecting the electron gun and any particular point on the screen, a probability amplitude $\mathscr{A}$(path). The net probability amplitude $K(2;1)$ for the particle to go from the event $\mathscr{P}1$ to the event $\mathscr{P}2$, is obtained by adding up the amplitudes for all the paths connecting the events:

$$K(2;1) \equiv K(t_2, x_2; t_1, x_1) = \sum_{\text{paths}} \mathscr{A}(\text{path}). \tag{2.3}$$

The addition of the *amplitudes* allows for the quantum mechanical interference between the paths. The *probability* for any process, of course, is obtained by taking the square of the modulus of the amplitude.

The quantity $K(t_2,x_2;t_2,x_1)$ contains the full dynamical information about the quantum mechanical system. Given $K(t_2,x_2;t_2,x_1)$ and the initial amplitude $\psi(t_2,x_1)$ for the particle to be found at x_1, we can compute the wave function $\psi(t,x)$ at any later time by the usual rules for combining the amplitudes:

$$\psi(t,x)=\int dx_1 K(t,x;t_1,x_1)\psi(t_1,x_1). \tag{2.4}$$

Therefore the specification of (i) $\mathscr{A}$ and (ii) the rule for evaluating the sum, in (2.3), will provide a complete quantum mechanical description of the system.

Since $K(t_2,x_1;t_1,x_1)$ contains the complete dynamical information of a quantum mechanical system, it is obvious that we will not be able to *derive* the rules for its computation from fundamental considerations. We have to prescribe a choice for $\mathscr{A}$ and the rule for computing the sum in (2.3). The usual choice for the amplitude $\mathscr{A}$ is

$$\mathscr{A}=\exp\left\{i\frac{A[x(t)]}{\hbar}\right\}. \tag{2.5}$$

Then the Kernel becomes

$$K(t_2,x_2;t_1,x_1)=\sum_{\text{all } x(t)}\exp i\frac{A[x(t)]}{\hbar}. \tag{2.6}$$

In the limit of $\hbar$ going to zero, the phase of $\mathscr{A}$ oscillates rapidly and the contributions from different paths are mostly cancelled out; the only ones that survive are those for which A is an extremum, viz. the classical paths. This choice in (2.6) thus provides a natural explanation for the validity of principle of least action in the classical theory.

The definition for "sum overpaths" is somewhat more complicated and depends on the form of $A[x(t)]$. For a wide class of lagrangians, this sum can be defined by a time-slicing method (see Feynman and Hibbs 1965; Shulman 1981; Rivers 1987). We shall, however, be concerned with a more restricted class of lagrangians which contain x and $\dot{x}$ only up to quadratic order. For these systems, the sum can be specified in a more useful manner, as a determinant of an operator. These are the systems described by an action of the form

$$A[x(t)]=\int(B(t)\dot{x}^2+C(t)x^2+M(t)x)dt. \tag{2.7}$$

A more general form for a quadratic action

$$A[x(t)]=\int[a(t)\dot{x}^2+b(t)\dot{x}+c(t)+x\dot{x}+d(t)x^2+e(t)x+f(t)]dt \tag{2.8}$$

can be reduced to the form in (2.7) by suitable partial integrations. It can be easily shown that the kernel in this case will be of the form

$$K(t_2,x_2;t_1,x_1)=N(t_1,t_2)\exp\left[\frac{i}{\hbar}A_c(t_2,x_2;t_1,x_1)\right] \tag{2.9}$$

where $N(t_2, t_1)$ stands for the path integral

$$N(t_2,t_1)=\sum_{\text{paths}}\exp\frac{i}{\hbar}\int_{t_1}^{t_2}\mathrm{d}t[B\dot{q}^2+Cq^2]$$

$$=\sum_{\text{paths}}\exp-\frac{i}{\hbar}\int_{t_1}^{t_2}\mathrm{d}tq\hat{D}q \tag{2.10}$$

with

$$\hat{D}=\left(\frac{\mathrm{d}}{\mathrm{d}t}B\frac{\mathrm{d}}{\mathrm{d}t}-C\right) \tag{2.11}$$

and the sum is over paths with the boundary condition $q(t_1)=q(t_2)=0$. We can now *define* the sum in such a way that

$$N(t_2,t_1)=(\det\hat{D})^{-1/2}. \tag{2.12}$$

We shall use the notation $\mathscr{D}q$ to indicate the sum over paths, when it is defined by the above prescription. Then

$$N=\int\mathscr{D}q\exp\left[-\frac{i}{\hbar}\int q\hat{D}q\mathrm{d}t\right]=(\det\hat{D})^{-\frac{1}{2}}. \tag{2.13}$$

We will adopt this notation when no confusion will arise.

The above prescription implicitly assumes that the operator $\hat{D}$ is treated as the limit of the expression:

$$\hat{D}\equiv\lim_{\varepsilon\to 0}(\hat{D}-i\varepsilon). \tag{2.14}$$

This procedure, called the '$i\varepsilon$-prescription' is just one of the many ways of making sense out of an ill-defined integral. It is possible to devise other modifications of the operator $\hat{D}$ – and corresponding limiting procedures – to give meaning to the integral. One such important alternative procedure, which is extensively used, is based on the method of analytically continuing the expressions into complex-t plane. Let us introduce a variable $\tau\equiv it$ (so that $t=-i\tau$) in the action. Under this substitution, the quantity

$$\exp\frac{i}{\hbar}Q=\exp\frac{i}{\hbar}\int_{t_1}^{t_2}\mathrm{d}t(B(t)\dot{q}^2+C(t)q^2) \tag{2.15}$$

becomes

$$\exp\left(\frac{i}{\hbar}Q\right)=\exp\left(-\frac{Q_E}{\hbar}\right)=\exp-\frac{1}{\hbar}\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\left(B_E(\tau)\left(\frac{\mathrm{d}q_E}{\mathrm{d}\tau}\right)^2-C_E(\tau)q_E^2\right) \tag{2.16}$$

where $B_E(\tau)\equiv B(t=-i\tau)$ etc. *We will assume that the original action is such that* (i) $B_E(\tau)$ *and* $C_E(\tau)$ *are real* (ii) $B_E(\tau)>0$ *and* (iii) $C_E(\tau)<0$. Then the argument of the exponent in (2.16) is negative definite for all real paths $q_E(\tau)$. [This set of paths, of course, is different from the set of paths obtained by substituting $t=-i\tau$ in the original set of paths; in general, if $q(t)$ is a real function, $q_E(\tau)=q(t=-i\tau)$ will not be real. In fact, one cannot even assume that a general path $q(t)$ can be analytically continued].

Let us now consider the sum

$$N_E(\tau_2,\tau_1)\equiv\sum_{\text{all real } q(\tau)}\exp-\frac{1}{\hbar}\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\left(B_E\left(\frac{\mathrm{d}q}{\mathrm{d}\tau}\right)^2-C_Eq^2\right)\tag{2.17}$$

which will be equal to $(\det D_E)^{-1/2}$ if we use the previous prescription. [The normalization in the prescription can be readjusted so that no extra i factors appear]. We can now *define* the original expression $N(t_2,t_1)$ as the analytic continuation of the quantity $N_E(\tau_2,\tau_1)$:

$$N(t_2,t_1)\equiv N_E(\tau_2=it_2;\tau_1=it_1).\tag{2.18}$$

This procedure may be summarized as follows: (i) From the original expression $Q[q(t)]$ obtain $Q_E[q(\tau)]$ by analytically continuing from t to τ. (ii) Check that B_E, C_E are real and $B_E>0$ and $C_E<0$. (iii) Evaluate the sum over paths Q_E, by summing over all real $q(\tau)$. (iv) Analytically continue back to t; this is *defined* to be the value of the original sum over paths. It should be emphasized that this method works only for those actions for which the condition (ii) above is satisfied. The quantity τ is called the 'Euclidean time' and other variables like A_E, G_E etc. are called 'Euclidean action', 'Euclidean Green function' etc. The two definitions for $N(t_2,t_1)$ given above will agree for a wide class of lagrangians, but not for *all* lagrangians.

2.2 *Kernels and ground-state expectation values*

We shall next discuss some relations between the Kernel and other quantities of interest. These relations, of course, are independent of the procedures used to calculate the Kernel; however, they are often used in combination with the path integral expression for the Kernel.

In the conventional approach to quantum mechanics, using the Heisenberg picture, the description of the system is in terms of the position and momentum operator $\hat{x}$ and $\hat{p}$. Let $|x,t\rangle$ be the eigenstate of the operator $\hat{x}(t)$ with eigenvalue x. The Kernel—which represents the probability amplitude for a particle to propagate from (t_1,x_1) to (t_2,x_2)—can be expressed, in a more conventional notation, as the matrix element:

$$K(t_2,x_2;t_1,x_1)=\langle x_2,t_2|t_1,x_1\rangle=\left\langle x_2,0\left|\exp\left\{-\frac{i}{\hbar}\hat{H}(t_2-t_1)\right\}\right|0,x_1\right\rangle\tag{2.19}$$

where $\hat{H}$ is the (time independent) Hamiltonian for the system. This relation allows one to represent the Kernel in terms of energy eigenstates of the system, provided the hamiltonian is independent of time. We have

$$\begin{aligned}K(T,x_2;0,x_1)&=\left\langle x_2,0\left|\exp\left(-\frac{i}{\hbar}HT\right)\right|0,x_1\right\rangle\\&=\sum_{n,m}\langle x_2|E_n\rangle\left\langle E_n\left|\exp\left(-\frac{i}{\hbar}HT\right)\right|E_m\right\rangle\langle E_m|x_1\rangle\\&=\sum_n\psi_n(x_2)\psi_n^*(x_1)\exp\left(-\frac{i}{\hbar}E_nT\right)\end{aligned}\tag{2.20}$$

where $\psi_n(x) = \langle x|E_n\rangle$ is the n-th energy eigenfunction of the system under consideration.
In physical applications, we often require the limiting form of

$$W(T;x_2,x_1) \equiv K(x_2t_2;x_1t_1) = K(x_2T;x_10). \quad (2.21)$$

for large T. This cannot be directly ascertained from (2.20) because the exponent oscillates. However, we can give meaning to this limit if we first transform (2.20) to the imaginary time $\tau_1 = it_1$ and $\tau_2 = it_2$ and consider form of W_E to large values of $(\tau_2 - \tau_1)$. We find that, in this limit,

$$W_E(T;x_2,x_1) \cong \psi_0(x_2)\psi_0(x_1)\exp\left[-\frac{E_0}{\hbar}(\tau_2-\tau_1)\right] \quad (2.22)$$

where the zero-subscript denotes the lowest energy state. (Note that $\psi_0^* = \psi_0$). From (2.22) we see that only the ground state contributes in this infinite time limit. We may now *define* the corresponding limit in (2.21) as the analytic continuation of (2.22), getting

$$W(T;x_2,x_1) \approx \psi_0(x_2)\psi_0(x_1)\exp\left(-i\frac{E_0T}{\hbar}\right). \quad (2.23)$$

This expression allows one to determine the ground state energy of the system from the Kernel in a simple manner. We see that

$$W_E(T;0,0) \approx (\text{constant})\exp\left(-\frac{E_0T}{\hbar}\right) \quad (2.24)$$

giving

$$E_0 = \lim_{T\to\infty}\left(-\frac{\hbar}{T}\ln W_E(T;0,0)\right). \quad (2.25)$$

The Kernel can also be used to study the effect of external perturbations on the system. Let us suppose that the system was in the ground state in the asymptotic past ($t_1 \approx -\infty$). At some time $t = -T$ we switch on an external time dependent disturbance $\lambda(t)$ affecting the system. Finally at $t = +T$ we switch off the perturbation. Because of the time-dependence, we no longer have stationary energy eigenstates for the system. In fact, the system is likely to have absorbed energy from the perturbation and would have ended up at some excited state at $t_2 = +\infty$; the probability for it to be found in the ground state as $t_2 = +\infty$ will be, in general, less than one. This probability can be computed from the Kernel. Consider the amplitude

$$\mathscr{P} = \lim_{t_2\to\infty}\lim_{t_2\to-\infty} K(t_2x_2;t_1x_1;\lambda(t)) = \lim_{t_2\to\infty}\lim_{t_1\to-\infty}\langle t_2x_2|t_1x_1\rangle_\lambda$$

$$= \lim_{t_2\to\infty}\lim_{t_1\to-\infty}\left[\int_{-\infty}^{+\infty} dx dx' \langle t_2,x_2|T,x\rangle\langle T,x|-T,x'\rangle_\lambda\langle -T,x'|t_1x_1\rangle\right]. \quad (2.26)$$

Since $\lambda = 0$ during $t_2 > t > T$ and $-T > t > t_1$, matrix elements in these intervals can be expressed in terms of the energy eigenstates of the original system for large t_2, $-t_1$:

$$\langle t_2,x_2|T,x\rangle \cong \psi_0(x_2)\psi_0(x)\exp\left[-i\frac{E_0}{\hbar}(t_2-T)\right]$$

$$\langle -T, x'|t_1, x_1\rangle \cong \psi_0(x')\psi_0(x_1)\exp\left[-i\frac{E_0}{\hbar}(-T-t_1)\right]. \tag{2.27}$$

Therefore (setting $\hbar = 1$ for simplicity), for large $(t_2 - t_1)$:

$$\begin{aligned}\mathscr{P} &\cong [\psi_0(x_2)\psi_0(x_1)\exp[-iE_0(t_2-t_1)]]\int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}x'(\psi_0(x)\exp(+iE_0T)) \\ &\quad\times \langle T, x|x', -T\rangle_\lambda(\psi_0(x')\exp(iE_0T)) \\ &\cong K(t_2, x_2; t_1 x_1; \lambda = 0)\int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}x'[\psi_0(x, T)]^* \\ &\quad\times \langle T, x|x', -T\rangle_\lambda[\psi_0(x', -T)]\end{aligned} \tag{2.28}$$

where $\psi_0(x, T)$ represents the ground state wave function at time T etc. The quantity

$$\mathscr{W} = \int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}x'[\psi_0(x, T)]^* \langle x, T|x', -T\rangle_\lambda[\psi_0(x', -T)] \tag{2.29}$$

represents the amplitude for the system to remain in the ground state in the asymptotic future if it started out in the ground state in the asymptotic past [usually called the "vacuum to vacuum" amplitude]. From (2.28) we find that this amplitude is given by the limit:

$$\mathscr{W} = \lim_{t_2\to\infty}\lim_{t_1\to-\infty}\frac{K(t_2, x_2; t_1, x_1; \lambda(t))}{K(t_2, x_2; t_1, x_1; 0)}. \tag{2.30}$$

This result can be further simplified by noticing that the x_2 and x_1 dependences cancel out in the ratio in (2.30) so that we can set $x_2 = x_1 = 0$, (or to any other constant value) getting

$$\mathscr{W} = \lim_{t_1\to\infty}\lim_{t_1\to-\infty}\frac{K(t_2, 0; t_1, 0; \lambda(t))}{K(t_2, 0; t_1, 0; 0)}. \tag{2.31}$$

Thus the vacuum-vacuum amplitude can be found from the Kernel by a simple limiting procedure.

This quantity $\mathscr{W}$ has a simple form when the external perturbation $\lambda(t)$ varies "adiabatically". That is, the perturbation $\lambda(t)$ varies slowly compared to the intrinsic time scales of the system. Then the ground state evolves in time adiabatically, as

$$\psi_0(x, t) = \psi_0(x, 0; \lambda)\exp -\frac{i}{\hbar}\int_0^t E_0(\lambda(t))\mathrm{d}t \tag{2.32}$$

where ψ_0 is the ground state and $E_0(\lambda)$ is the ground state energy of the hamiltonian calculated by treating λ as some given, time-independent parameter. In this case, it is easy to see that

$$\mathscr{W} = \exp -\frac{i}{\hbar}\int_{-\infty}^{+\infty} E_0(\lambda(t))\mathrm{d}t. \tag{2.33}$$

We will use this result later.

3. The effective action

3.1 *The concept of effective action*

Consider a theory which describes the interaction between two systems having the dynamical variables Q and q. [This notation is purely formal; the symbol Q, for example, could describe a *set* of variables, like the components of a vector field. The detailed nature of these variables is not of importance at this stage.] The full quantum theory can be constructed from the exact Kernel

$$K(Q_2,q_2;Q_1,q_1;t_2,t_1)=\int \mathscr{D}Q\int \mathscr{D}q \exp\frac{i}{\hbar}A[Q,q] \tag{3.1}$$

which is often impossible to evaluate. It would be, therefore, useful to have some approximate ways of studying the system.

The 'effective action' method is *one* of the *many* approximation schemes available for handling (3.1). This method is of value when one of the variables, say, Q, behaves nearly classically while the other variable is fully quantum mechanical. In that case, the problem can be attacked in the following manner:

Let us suppose that the path integral over q in (3.1) can be performed exactly, for an arbitrary $Q(t)$. That is, we can evaluate the quantity

$$F[Q(t);q_2,q_1;t_2,t_1]\equiv \exp\frac{i}{\hbar}W=\int \mathscr{D}q \exp\left(\frac{i}{\hbar}A[Q(t),q]\right) \tag{3.2}$$

treating $Q(t)$ as any specified function of time. If we could now do

$$K=\int \mathscr{D}Q \exp\left(\frac{i}{\hbar}W[Q]\right) \tag{3.3}$$

exactly, we would have completely solved the problem. Since this is not possible, we will evaluate (3.3) by invoking the fact that Q is almost classical. This means that most of the contribution to (3.3) comes from nearly classical paths satisfying the condition

$$\frac{\delta W}{\delta Q}=0. \tag{3.4}$$

It is usually easy to evaluate (3.3) in this approximation and thereby obtain an approximate solution to our problem. In fact, quite often, we will be content with obtaining the solutions to (3.4), and will not even bother to calculate (3.3) in this approximation. Equation (3.4), of course, will contain some of the effects of the quantum fluctuations of q on Q, and is often called the 'semiclassical equation'. The quantity W is called the 'effective action' for the Q-system. In general, the functional $W[Q(t)]$ cannot be expressed as an integral over time of a local density. Whenever it is possible, we can define an 'effective lagrangian' through the relation

$$W=\int L_{\text{eff}}\,dt. \tag{3.5}$$

The way we have defined our expressions, the quantities K and W depend on the boundary conditions (t_2,q_2,t_1,q_1). It is preferable to have an effective action which is completely independent of the q-degree of freedom. The most natural way of achieving this is to integrate out the effect of q for *all times* by considering the limit $t_2 \to +\infty$, $t_1 \to -\infty$ in our definition of the effective action. We will also assume, as is usual, that $Q(t)$ becomes constant asymptotically. From our discussion in part 2 we know that, in this limit, the Kernel essentially represents the amplitude for the q-system to make a transition from the ground state in infinite past to the ground state in the infinite future. Hence

$$F(q_2, q_1; +\infty, -\infty) \equiv \exp \frac{i}{\hbar} W[Q(t)] = N(q_2, q_1) \langle E_0, +\infty | -\infty, E_0 \rangle_{Q(t)} \tag{3.6}$$

where $\langle E_0, +\infty | -\infty, E_0 \rangle_{Q(t)}$ stands for the 'vacuum to vacuum' amplitude for the q-system in the presence of the external source $Q(t)$ and $N(q, q_2)$ is a normalization factor, independent of $Q(t)$. Taking logarithms we get

$$W[Q(t)] = -i\hbar \ln \langle E_0, +\infty | E_0, -\infty, \rangle_{Q(t)} + (\text{constant}). \tag{3.7}$$

Since the constant term is independent of Q it will not contribute in (3.4). Therefore, for the purposes of our calculation we may take the effective action to be defined by the relation

$$W[Q(t)] \equiv i\hbar \ln \langle E_0, +\infty | E_0, -\infty, \rangle_{Q(t)} \tag{3.8}$$

in which all reference to the quantum mode q is eliminated. Notice that the way we have defined our F, the effective action W contains the kinetic energy of Q and any potential energy of Q [which depends only on Q]. That is, if the original lagrangian has the form $L = (1/2)\dot{Q}^2 - V(Q) + (1/2)\dot{q}^2 - u(Q, q)$, the effective action will have the form $W = (1/2)\dot{Q}^2 - V(Q) + W_c[Q]$; the first two terms of L go for a ride and the last term W_c is the result of integrating out q.

This discussion also highlights an important feature of the effective action. We have seen in part 2 that an external perturbation can cause transitions in a system from ground state to excited state. In other words, the probability for the system to be in the ground state in the infinite future (even though it started in the ground state in the infinite past) could be less than unity. This implies that our effective action W_0 need not be real. If we use this W directly in (3.4) we have no assurance that our solution Q will be real. In fact, the saddle point approximation has to be handled with care if W is complex. The imaginary part of W contains information about the rate of transitions induced in the q-system by the presence of $Q(t)$; or—in the context of field theory—the rate of production of particles from the vacuum. The semiclassical equation is of very doubtful validity if these excitations drain away too much energy from the Q-mode. Thus we must confine ourselves to the situations in which

$$\text{Im}\, W \ll \text{Re}\, W. \tag{3.9}$$

In that case, we can modify the semiclassical equations to read

$$\frac{\delta \text{Re}\, W}{\delta Q} = 0. \tag{3.10}$$

Usually, the action $A[Q, q]$ will have the form $A_0[Q] + A_I(Q, q)$ where A_0 is the 'free' part and A_I represents the interaction between Q and q. Then W can be expressed as $(A_0 + W_{corr})$ with a real A_0. The condition for the suppression of particle production now becomes $(A_0 + \text{Re}\, W_c) \gg \text{Im}\, W_c$. This can be satisfied even if $\text{Re}\, W_c \simeq \text{Im}\, W_c$, as long as A_0 is large compared to $|W_c|$.

In most practical situations, the constraint (3.9) will automatically arise because of another reason. Notice that the entire scheme depends on our ability to evaluate the first path integral in (3.2). This task is far from easy, especially because this expression is needed for an arbitrary $Q(t)$. Quite often, one evaluates this expression by assuming that the time variation of $Q(t)$ is slow compared to time scales over which the quantum variable q fluctuates. In such a case, the characteristic frequencies of the q-mode will be much higher than the frequency at which Q-mode is evolving and hence there will be very little transfer of energy from Q to q. The real part of W will dominate.

The above discussion allows an alternative picture of the effective action which is quite useful. Let us suppose that $Q(t)$ varies slowly enough for the adiabatic approximation to be valid for q. We then know—from our discussion in part 2—that the 'vacuum to vacuum' amplitude is given by:

$$\lim_{t_2 \to \infty} \lim_{t_1 \to -\infty} F(q_2, q_1; t_2, t_1) = \mathcal{W} = (\text{constant}).\exp\left(-\frac{i}{\hbar} \int_{-\infty}^{+\infty} E_0(Q)\, dt \right). \tag{3.11}$$

This expression allows us to identify the effective lagrangian as the ground state energy of the q-mode in the presence of Q:

$$L_{\text{eff}} = -E_0(Q). \tag{3.12}$$

This result, which is valid when the time dependence of Q is treated in the adiabatic limit in the calculation of E_0, provides an alternative means of computation of the effective lagrangian if the Q dependence of the ground state energy can be ascertained.

The transitions to the higher states, indicated by the existence of an imaginary part to W_0, can also be discussed in terms of the above relation. The W_0 can become complex only if L_{eff} and hence E_0 becomes complex. The appearance of an imaginary part to the ground state energy indicates an exponential decay probability for this state with some half life. This is precisely what we expect if transitions to higher states are possible.

The above discussion may suggest that whenever Q varies slowly enough the real part of W—or, equivalently, the real part of L_{eff}—will give the dominant contribution. If that is the case, we should get no imaginary part to W when the time variation of Q is highly suppressed by treating Q as an adiabatically varying parameter. This is *usually* true but one must make sure that a ground state *exists* for the range of Q values considered in the problem. As a simple example, consider the action

$$A[Q, q] = \int dt\left(\tfrac{1}{2}\dot{Q}^2 + \tfrac{1}{2}\dot{q}^2 - \tfrac{1}{2}(\omega^2 - Q^2)q^2\right). \tag{3.13}$$

We see that q behaves as a harmonic oscillator with the effective frequency

$$\omega_{\text{eff}} = \sqrt{\omega^2 - Q^2}. \tag{3.14}$$

It is possible to arrange matters so that Q becomes larger than ω in the course of the evolution even though Q vanished in the asymptotic past and was increasing arbitrarily slowly. If this happens, no vacuum state will exist for the q-mode for a certain range of Q and our calculation will lead to an imaginary part for the effective action. So, in general, the existence of an imaginary part to the effective action may either be due to transitions to higher states or due to the non-existence of the ground state. [If we interpret adiabatically as smallness of $(\dot{\omega}_{\rm eff}/\omega^2_{\rm eff})$, then adiabaticity is violated when $\omega_{\rm eff}$ vanishes]. In the course of our discussion we will come across examples of both situations.

3.2 *The method of proper time*

The expressions for the effective action simplify considerably, when the q-dependence of the action is quadratic, or can be approximated as quadratic. Consider, for example, the system with two scalar fields $\Phi_c(x)$ and $\phi(x)$ with the lagrangian

$$L_{\rm total} = L_0(\Phi_c) + \tfrac{1}{2}(\phi^i\phi_i - m^2(\Phi_c)\phi^2) = L(\Phi_c) + L_{\rm int} \tag{3.15}$$

where $m^2(\Phi_c)$ is some function of Φ_c. The correction term $L_{\rm int}$ represents a scalar field with effective mass $m^2(\Phi_c)$. We will treat ϕ as a quantum variable and Φ_c as a classical variable and are interested in the effect of quantum fluctuation in ϕ on Φ_c. In the adiabatic limit, in which Φ is varying sufficiently slowly, the effective lagrangian and the potential are given by

$$L_{\rm eff} = L_0 - E_0(m^2);\ V_{\rm eff} = V + E_0(m^2) \tag{3.16}$$

where $E_0(m^2)$ is the ground state energy of a scalar field theory with mass m which can be written as

$$E_0 = \frac{1}{2}\hbar \int \frac{{\rm d}^D k}{(2\pi)^D}(k^2 + m^2)^{1/2} \tag{3.17}$$

[For the sake of generality, we are considering a spacetime of $(D+1)$ dimensions]. Since this expression is badly divergent, we need to consider methods for making sense out of this expression. We will address ourselves to this question of 'renormalization' later. Before that, we will first cast this expression in a more manageable form.

It is convenient at this stage to introduce the Euclidean continuation. Since the energy in the Euclidean sector differs by a sign from that in Lorentzian space we need to calculate

$$L_{\rm corr}^{\rm (Euclidean)} = +E_0(m^2) = \frac{1}{2}\int \frac{{\rm d}^D k}{(2\pi)^D}(\mathbf{k}^2 + m^2)^{1/2} \equiv L_c \tag{3.18}$$

where $\mathbf{k}$ is a D-dimensional vector. Note that, we can write

$$\begin{aligned}\frac{\partial L_c}{\partial m^2} &= \frac{1}{4}\int \frac{{\rm d}^D k}{(2\pi)^D}\frac{1}{(\mathbf{k}^2 + m^2)^{1/2}} \\ &= \frac{1}{4}\int \frac{{\rm d}^D k}{(2\pi)^D}\cdot\frac{2}{\sqrt{\pi}}\int_0^\infty {\rm d}\lambda \exp(-\lambda^2(\mathbf{k}^2 + m^2)) \\ &= \frac{1}{4}\int \frac{{\rm d}^D k}{(2\pi)^D}\int_0^\infty \frac{{\rm d}s}{(2\pi s)^{1/2}}\exp(-\tfrac{1}{2}s(k^2 + m^2)).\end{aligned} \tag{3.19}$$

The $s^{-1/2}$ factor can be eliminated by the following trick. We introduce a variable p and rewrite this factor as another integral

$$\frac{1}{(2\pi s)^{1/2}}=\int_{-\infty}^{+\infty}\frac{\mathrm{d}p}{2\pi}\exp(-\tfrac{1}{2}sp^2) \tag{3.20}$$

obtaining

$$\frac{\partial L_c}{\partial m^2}=\frac{1}{4}\int\frac{\mathrm{d}^Dk}{(2\pi)^D}\int_0^\infty \mathrm{d}s\int_{-\infty}^{+\infty}\frac{\mathrm{d}p}{2\pi}\exp[-\tfrac{1}{2}s(\mathbf{k}^2+p^2+m^2)]. \tag{3.21}$$

The $\mathbf{k}$ and p integrations can be combined into a $(D+1)$-dimensional integration over the vector $\mathbf{q}=(\mathbf{k},p)$:

$$\begin{aligned}\frac{\partial L_c}{\partial m^2}&=\frac{1}{4}\int\frac{\mathrm{d}^{D+1}q}{(2\pi)^{D+1}}\int_0^\infty \mathrm{d}s\exp[-\tfrac{1}{2}s(\mathbf{q}^2+m^2)]\\ &=\frac{1}{4}\int_0^\infty \mathrm{d}s\exp[-\tfrac{1}{2}sm^2]\int\frac{\mathrm{d}^{D+1}q}{(2\pi)^{D+1}}\exp[-\tfrac{1}{2}s\mathbf{q}^2]\\ &=\frac{1}{4}\int_0^\infty\frac{\mathrm{d}s}{(2\pi s)^{(D+1)/2}}\exp[-\tfrac{1}{2}m^2 s].\end{aligned} \tag{3.22}$$

[Alternatively, one can do the d^Dk integration in (3.19)) to obtain this result.] Integrating this expression with respect to m^2, we get

$$L_c=-\frac{1}{2}\int_0^\infty\frac{\mathrm{d}s}{s(2\pi s)^{(D+1)/2}}\exp[-\tfrac{1}{2}m^2 s] \tag{3.23}$$

where we have omitted an integration constant which is independent of m^2. As it stands (3.23) is also divergent at $s=0$; however, in this form the divergences are easy to isolate and handle. Some of the manipulations above are not valid for integrals which are divergent. It is tacitly assumed that the integrals can be expressed as limits of some well-defined convergent integrals.

There is another way of deriving (3.23) which is more straightforward (though it hides the physical meaning of L_{eff}) and is quite useful. The effect of quantum fluctuations ϕ is contained in the Kernel

$$K\equiv\exp\left(-\int \mathrm{d}x_E L_{\text{corr}}\right)=\int\mathscr{D}\phi\exp\left(-\int\mathrm{d}x\phi\hat{D}\phi\right)=(\det\hat{D})^{-1/2} \tag{3.24}$$

where D is the Euclidean space operator

$$\hat{D}=-\tfrac{1}{2}(\Box-m^2) \tag{3.25}$$

with $\Box$ denoting the $(D+1)$ dimensional Dalembertian. [containing D-space and 1 Euclidean time]. We will now write this determinant as

$$\det D=\exp[\operatorname{Tr}\ln D] \tag{3.26}$$

so that the Kernel becomes

$$(\det D)^{-1/2}=\exp(-\tfrac{1}{2}\operatorname{Tr}\ln D)=\exp\left(-\tfrac{1}{2}\int\mathrm{d}x\langle x|\ln D|x\rangle\right)\equiv\exp\left(-\int\mathrm{d}x\,L_{\text{corr}}\right). \tag{3.27}$$

In arriving at the last expression, we have used some basis vectors $|x\rangle$ to evaluate the trace. We will now use the integral representation for the logarithm,

$$\ln F = -\int_0^\infty \frac{ds}{s}\exp(-Fs) \tag{3.28}$$

to get

$$L_{corr} = \tfrac{1}{2}\langle x|\ln D|x\rangle = -\frac{1}{2}\int_0^\infty \frac{ds}{s}\langle x|\exp-(sD)|x\rangle$$

$$= -\frac{1}{2}\int_0^\infty \frac{ds}{s} K(x,x;s) \tag{3.29}$$

where the quantity

$$K(x,y;s) = \langle x|\exp(-sD)|y\rangle \tag{3.30}$$

is the Euclidean Kernel for a *quantum-mechanical* particle with the hamiltonian D. The integral representation given above is divergent at $s=0$. However, this expression can be used to study *difference* between two logarithms; we shall use this only in the latter sense.

This result is of very general validity and quite powerful (Schwinger 1951). It shows that if the Euclidean action coupling two systems has the form

$$A_{corr}[\Phi,\phi] = \int \phi \hat{D}_\Phi \phi \, d^{D+1}x \tag{3.31}$$

where $\hat{D}_\Phi$ is an operator depending on Φ, then the correction term in effective lagrangian is given by

$$L_{corr} = -\frac{1}{2}\int_0^\infty \frac{ds}{s} K(x,x;s) \tag{3.32}$$

where $K(x,y;s)$ represents the propagation Kernel for some fictitious quantum *mechanical* particle described by the hamiltonian in $(D+1)$ dimension

$$\hat{h} = \hat{D}_\Phi. \tag{3.33}$$

In other words, we have reduced the problem involving a path integral over *fields* to a problem involving quantum *mechanical* Kernel. The latter is often much easier to evaluate. The example we are concerned with has the hamiltonian

$$h = D = \tfrac{1}{2}(-\Box + m^2) = -\frac{1}{2}\left(\frac{d^2}{d\tau^2} + \nabla^2\right) + \tfrac{1}{2}m^2. \tag{3.34}$$

The lagrangian corresponding to this hamiltonian is

$$l = +\frac{1}{2}\left(\left(\frac{d\tau}{ds}\right)^2 + \left|\frac{d\mathbf{x}}{ds}\right|^2\right) - \tfrac{1}{2}m^2 \tag{3.35}$$

which represents a free particle in $(D+1)$ dimensional space with a constant background potential $(m^2/2)$. [Note that m^2 is treated as a constant in the adiabatic limit]. The Kernel K we need is that of a free particle:

$$K(x,x;s)=\left(\frac{1}{2\pi s}\right)^{(D+1)/2}\exp(-\tfrac{1}{2}m^2 s). \tag{3.36}$$

We thus get the expression for the effective lagrangian to be

$$L_{\text{eff}}=-\frac{1}{2}\int_0^\infty\frac{ds}{s}\left(\frac{1}{2\pi s}\right)^{(D+1)/2}\exp(-\tfrac{1}{2}m^2 s) \tag{3.37}$$

which agrees with the previous result. [Note the way in which i-factors disappeared in the Kernel. In arriving at the last two expressions we have proceeded as follows: The quantity $\langle x'|\exp - iTD|x\rangle$ with $D=-\frac{1}{2}\Box+\frac{1}{2}m^2$ is a proper Schrödinger Kernel with D as hamiltonian and T as time. Therefore

$$K(T;x,x)=\left(\frac{1}{2\pi iT}\right)^{(D+1)/2}\exp\left(-\frac{i}{2}m^2 T\right) \tag{3.38}$$

changing to $iT=s$ leads to the expression given above].

The Kernel $K(x,y;s)$ can also be used to compute another important quantity, viz. the propagator. Since the propagator G is the inverse of the operator $\hat{D}_\Phi$, it follows that

$$G(x,x')=\int_0^\infty ds K(x,x';s). \tag{3.39}$$

4. Quantum theory in external electromagnetic field

The formalism developed in the previous sections can now be applied to the study of an important problem: The calculation of the effective action for electromagnetic fields which will allow us—for example—to determine the quantum corrections to classical Maxwell equations. The study also reveals several important conceptual issues in our formalism. As a bonus we will be able to understand some aspects of the renormalization procedure in quantum electrodynamics.

4.1 *Effective action from ground state energy*

Consider a system described by the lagrangian density $L(A_i,\phi)$ where $A_i(x)$ is a vector potential describing the electromagnetic field and ϕ is a charged (complex) scalar field interacting with the electromagnetic field. The full quantum theory is described by the Kernel

$$K=\int\mathscr{D}A_i\mathscr{D}\phi\exp\left[i\int L dt\, dx\right] \tag{4.1}$$

in which we have set $\hbar=1$ for convenience. The effective action A_{eff} (and the effective lagrangian L_{eff}) for electrodyanmics can be obtained by integrating over the scalar field:

$$\exp(iA_{\text{eff}}) = \exp\left[i\int \mathrm{d}t\,\mathrm{d}\mathbf{x}\,L_{\text{eff}}(A_i)\right]$$
$$= \int \mathscr{D}\phi \exp\left[i\int \mathrm{d}t\,\mathrm{d}\mathbf{x}\,L(A_i,\phi)\right]. \tag{4.2}$$

Thus we need to evaluate the path integral over ϕ in a given background electromagnetic field.

As usual, this is an impossibly difficult task if $A_i(x)$ is an arbitrary background field. To make progress we will assume that $A_i(x)$ varies slowly with x so that we can write

$$A_i(x) \cong -\tfrac{1}{2}F_{ik}x^k + O((\partial F)x^2) \tag{4.3}$$

where F_{ik} are treated as constant. This corresponds to assuming that the background potential describes a constant electromagnetic field F_{ik}, or—more precisely—the field ϕ varies much more rapidly compared to the background electromagnetic field. Thus we will compute, in the adiabatic approximation:

$$\exp[iA_{\text{eff}}(F)] = \exp\left[i\int \mathrm{d}t\,\mathrm{d}\mathbf{x}\,L_{\text{eff}}(F)\right]$$
$$= \int \mathscr{D}\phi \exp\left[i\int \mathrm{d}t\,\mathrm{d}\mathbf{x}\,L[A_i = -1/2F_{ik}x^k,\phi]\right]. \tag{4.4}$$

We have seen earlier that, in the adiabatic limit we are considering, L_{eff} is the negative of the ground state energy of the system. Thus if we compute the ground state energy $E_0(F)$ of a scalar field ϕ in a given background F_{ik}, then we can determine $L_{\text{eff}}(F) = -E_0(F)$.

This task is particularly easy if the background field satisfies the conditions $\mathbf{E}\cdot\mathbf{B} = 0$ and $\mathbf{B}^2 - \mathbf{E}^2 > 0$. (This derivation is adapted from Berestetskii *et al* (1979)). In such a case, the field can be expressed as purely magnetic in some Lorentz frame. Let $\mathbf{B} = (0, B, 0)$; we choose the gauge such that $A^i = (0,0,0,-Bx)$. The Klein-Gordon equation

$$[(i\partial_\mu - qA_\mu)^2 - m^2]\phi = 0 \tag{4.5}$$

can now be separated by taking

$$\phi(t,\mathbf{x}) = f(x)\exp i(k_y y + k_z z - \omega t). \tag{4.6}$$

where $f(x)$ satisfies the equation

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} + [\omega^2 - (qBx - k_z)^2]f = (m^2 + k_y^2)f. \tag{4.7}$$

This can be rewritten as

$$-\frac{\mathrm{d}^2 f}{\mathrm{d}\xi^2} + q^2B^2\xi^2 f = \varepsilon f \tag{4.8}$$

where

$$\xi = x - \frac{k_z}{qB}; \quad \varepsilon = \omega^2 - m^2 - k_y^2. \tag{4.9}$$

Equation (4.8) is that of a harmonic oscillator with mass (1/2) and frequency $2(qB)$. So, if $f(x)$ has to be bounded for large x, the energy ε must be quantized:

$$\varepsilon_n = 2(qB)(n + \tfrac{1}{2}) = \omega^2 - (m^2 + k_y^2). \tag{4.10}$$

Therefore the allowed set of frequencies is

$$\omega_n = [m^2 + k_y^2 + 2qB(n + \tfrac{1}{2})]^{1/2}. \tag{4.11}$$

The ground state energy per mode is $2(\omega_n/2) = \omega_n$ because the complex scalar field has twice as many degrees of freedom as a real scalar field. The total ground state energy is given by the sum over all modes k_y and n. The weightage factor for the discrete sum over n, in a magnetic field is obtained by the correspondence:

$$\frac{dk_x}{2\pi}\frac{dk_y}{2\pi} \to \sum_n \left(\frac{qB}{2\pi}\right)\frac{dk_y}{2\pi}. \tag{4.12}$$

Hence, the ground state energy is

$$E_0 = \sum_{n=0}^{\infty}\left(\frac{qB}{2\pi}\right)\int_{-\infty}^{+\infty}\frac{dk_y}{(2\pi)}\left[(k_y^2 + m^2) + 2qB\left(n + \frac{1}{2}\right)\right]^{1/2} = -L_{\text{eff}}. \tag{4.13}$$

This expression, as usual, is divergent. To separate out a finite part we will proceed as follows: Consider the quantity

$$I \equiv -\left(\frac{2\pi}{qB}\right)\frac{\partial^2 E_0}{\partial(m^2)^2} = \left(\frac{2\pi}{qB}\right)\frac{\partial^2 L_{\text{eff}}}{\partial(m^2)^2}. \tag{4.14}$$

which can be evaluated in the following manner:

$$\begin{aligned}
I &= +\frac{1}{4}\sum_{n=0}^{\infty}\int_{-\infty}^{+\infty}\frac{1}{2\pi}\frac{dk_y}{[k_y^2 + m^2 + 2qB(n + \frac{1}{2})]^{3/2}} = +\frac{1}{8\pi}\sum_{n=0}^{\infty}\frac{1}{[m^2 + 2qB(n + \frac{1}{2})]}\cdot 2 \\
&= -\frac{1}{4\pi}\sum_{n=0}^{\infty}\int_0^{\infty} d\eta \exp[-\eta(m^2 + 2qB(n + \tfrac{1}{2}))] \\
&= +\frac{1}{4\pi}\int_0^{\infty} d\eta \exp(-\eta m^2)\cdot\exp(-qB\eta)\cdot\frac{1}{1 - \exp(-2qB\eta)} \\
&= +\frac{1}{4\pi}\int_0^{\infty} d\eta \frac{\exp(-\eta m^2)}{\exp(qB\eta) - \exp(-qB\eta)} = \frac{1}{8\pi}\int_0^{\infty} d\eta \frac{\exp(-\eta m^2)}{\sinh qB\eta} = \left(\frac{2\pi}{qB}\right)\frac{\partial^2 L_{\text{eff}}}{\partial(m^2)^2}.
\end{aligned} \tag{4.15}$$

The L_{eff} can be determined by integrating the expression twice with respect to m^2. We get

$$\begin{aligned}
L_{\text{eff}} &= \frac{qB}{(4\pi)^2}\int_0^{\infty}\frac{d\eta}{\eta^2}\frac{\exp(-\eta m^2)}{\sinh qB\eta} \\
&= \int_0^{\infty}\frac{d\eta}{(4\pi)^2}\cdot\frac{\exp(-\eta m^2)}{\eta^3}\cdot\frac{qB\eta}{\sinh qB\eta}.
\end{aligned} \tag{4.16}$$

This—and the subsequent expressions—has a divergence at the lower limit of integration. This divergence can be removed by subtracting the contribution with $E=B=0$; we will ignore this problem right now and will take it up later in §4·3. The integration with respect to m^2 also produces a term like $(c_1 m^2 + c_2)$ with two (divergent) integration constants c_1 and c_2. We have not displayed this term here; this divergence is also connected with the "renormalization" of L_{eff} and will be discussed later.

If the L_{eff} has to be Lorentz and gauge invariant then it can only depend on the quantities $(E^2 - B^2)$ and $\mathbf{E}\cdot\mathbf{B}$. We will define two constants a and b by the relation

$$a^2 - b^2 = E^2 - B^2; \; ab = \mathbf{E}\cdot\mathbf{B}. \tag{4.17}$$

Then $L_{\text{eff}} = L_{\text{eff}}(a, b)$. In the case of pure magnetic field we are considering $a = 0$ and $b = B$. Therefore, the L_{eff} can be written in a manifestedly invariant way as:

$$L_{\text{eff}} = \int_0^\infty \frac{d\eta}{(4\pi)^2} \cdot \frac{\exp(-\eta m^2)}{\eta^3} \cdot \frac{qb\eta}{\sinh qb\eta}. \tag{4.18}$$

Because this form is Lorentz invariant, it must be valid in any frame in which $E^2 - B^2 < 0$ and $\mathbf{E}\cdot\mathbf{B} = 0$. In all such cases,

$$L_{\text{eff}} = \int_0^\infty \frac{d\eta}{(4\pi)^2} \frac{\exp(-\eta m^2)}{\eta^3} \frac{q\eta\sqrt{B^2 - E^2}}{\sinh q\eta\sqrt{B^2 - E^2}}. \tag{4.19}$$

The L_{eff} for a pure electric field can be determined from this expression if we analytically continue the expression even for $B^2 < E^2$. We will find, for $B = 0$,

$$L_{\text{eff}} = \int_0^\infty \frac{d\eta}{(4\pi)^2} \frac{\exp(-\eta m^2)}{\eta^3} \frac{q\eta E}{\sin q\eta E}. \tag{4.20}$$

The same result can be obtained by noticing that a and b are invariant under the transformation $E \to iB, B \to iE$. Therefore, $L_{\text{eff}}(a, b)$ must also be invariant under these transformations: $L_{\text{eff}}(E, B) = L_{\text{eff}}(iB, -iE)$. This allows us to get (4.20) for (4.16).

We will now consider the general case with arbitrary $\mathbf{E}$ and $\mathbf{B}$ for which a and b are not simultaneously zero. It is well-known that by choosing our Lorentz frame suitably, we can make E and B parallel, say along the y-axis. We will describe this field $[\mathbf{E} = (0, E, 0); \mathbf{B} = (0, B, 0)]$ in the gauge $A_i = [-Ey, 0, 0, -Bx]$. The Klein-Gordon equation becomes

$$[(i\partial_\mu - qA_\mu)^2 - m^2]\phi = \left[\left(i\frac{\partial}{\partial t} + qEy\right)^2 + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \left(i\frac{\partial}{\partial z} + qBx\right)^2 - m^2\right]\phi = 0. \tag{4.21}$$

Separating the variables by assuming

$$\phi(t, \mathbf{x}) = f(x, y)\exp - i(\omega t - k_z z) \tag{4.22}$$

we get

$$\left[\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + (\omega + qEy)^2 - (k_z - qBx)^2\right] f = m^2 f \tag{4.23}$$

which separates out into x and y modes. Writing

$$f(x, y) = g(x)Q(y) \tag{4.24}$$

where $g(x)$ satisfies the harmonic oscillator equation

$$\frac{d^2 g}{dx^2} - (k_z - qBx)^2 g = -2qB\left(n + \frac{1}{2}\right)g \tag{4.25}$$

we get

$$\frac{d^2 Q}{dy^2} + (\omega + qEy)^2 Q = \left[m^2 + 2qB\left(n + \frac{1}{2}\right)\right]Q. \tag{4.26}$$

Changing to the dimensionless variable

$$\eta = y\sqrt{qE} + \frac{\omega}{\sqrt{qE}} \tag{4.27}$$

we obtain

$$\frac{d^2 Q}{d\eta^2} + \eta^2 Q = \frac{1}{qE}\left(m^2 + 2qB\left(n + \frac{1}{2}\right)\right)Q. \tag{4.28}$$

To proceed further, we use a trick due to Landau. The expression shows that the only dimensionless combination which appears in the presence of an electric field is $\tau = (qE)^{-1}(m^2 + qB(2n+1))$. Thus, purely from dimensional considerations, we expect the ground state energy to have the form

$$E_0 = \sum_{n=0}^{\infty} (2qB)G(\tau) \tag{4.29}$$

where G is a function to be determined. Introducing the Laplace transform F of G, by the relation

$$G(\tau) = \int_0^{\infty} F(k)\exp[-k\tau]dk \tag{4.30}$$

we can write

$$L_{\text{eff}} = (2qB)\sum_{n=0}^{\infty}\int_0^{\infty} dk F(k)\exp\left[-\frac{k}{qE}(m^2 + qB(2n+1))\right]. \tag{4.31}$$

Summing the geometric series, we obtain

$$\begin{aligned} L_{\text{eff}} &= 2(qB)(qE)\int_0^{\infty} ds F(qEs)\exp[-sm^2]\exp[-qBs]\cdot\frac{1}{1-\exp[-2qBs]} \\ &= 2(qB)(qE)\int_0^{\infty} ds\frac{F(qEs)\exp[-sm^2]}{\exp[qBs] - \exp[-qBs]} \\ &= (qB)(qE)\int_0^{\infty} ds\frac{F(qEs)}{\sinh qBs}\exp[-m^2 s]. \end{aligned}$$

We now determine F by using the fact that L_{eff} must be invariant under the transformation $E \to iB, B \to -iE$. This means that

$$L_{\text{eff}} = (qB)(qE) \int_0^\infty \mathrm{d}s \exp[-m^2 s] \frac{F(iqBs)}{-\sinh(iqEs)}.$$

Comparing the two expressions and using the uniqueness of the Laplace transform with respect to m^2, we get

$$\frac{F(qEs)}{\sinh qBs} = -\frac{F(iqBs)}{\sinh(iqEs)} \tag{4.32}$$

or, equivalently,

$$F(qEs)\sin qEs = F(iqBs)\sin(iqBs). \tag{4.33}$$

Since each side depends only on either E or B alone, each side must be independent of E and B. Therefore

$$F(qEs)\sin qEs = F(iqBs)\sin(iqBs) = \text{constant} = A(s) \tag{4.34}$$

giving

$$L_{\text{eff}} = (qB)(qE) \int_0^\infty \mathrm{d}s \frac{\exp[-m^2 s] A(s)}{\sin qEs \sinh qBs}. \tag{4.35}$$

The $A(s)$ can be determined by comparing this expression with, say, (4.16) in the limit of $E \to 0$. We have

$$L_{\text{eff}}(E=0, B) = qB \int_0^\infty \frac{\mathrm{d}s}{s} \exp[-m^2 s] \cdot \frac{A(s)}{\sinh qBs}$$

$$= qB \int_0^\infty \frac{\mathrm{d}s}{(4\pi)^2 s^2} \exp[-m^2 s] \frac{1}{\sinh qBs} \tag{4.36}$$

implying

$$A(s) = \frac{1}{(4\pi)^2 s}. \tag{4.37}$$

Thus we arrive at the final answer

$$L_{\text{eff}} = \int_0^\infty \frac{\mathrm{d}s}{(4\pi)^2} \frac{\exp[-m^2 s]}{s^3} \left(\frac{qEs}{\sin qEs}\right)\left(\frac{qBs}{\sinh qBs}\right). \tag{4.38}$$

In the situation we are considering **E** and **B** are parallel making $a^2 - b^2 = E^2 - B^2$ and $ab = \mathbf{E}\cdot\mathbf{B}. = EB$. Therefore $E = a$ and $B = b$. Thus our result can be written in a manifestedly invariant form as

$$L_{\text{eff}}(a, b) = \int_0^\infty \frac{\mathrm{d}s}{(4\pi)^2} \frac{\exp[-m^2 s]}{s^3} \left(\frac{qas}{\sin qas}\right)\left(\frac{qbs}{\sinh qbs}\right). \tag{4.39}$$

This result will be now valid in any gauge or frames with a and b determined in terms of $(\mathbf{E}^2 - \mathbf{B}^2)$ and $(\mathbf{E}\cdot\mathbf{B})$.

The integral, as it stands, is ill-defined for two different reasons. (i) The sine function has poles along the path of integration at $qas = n\pi; n = 1,2,\ldots$ (ii) The integral diverges at $s = 0$. The second problem is related to renormalization and will be taken up in the next section while the first problem can be tackled in the following way:

The integral is evaluated by going around each of the poles by a small semicircle in the upper half plane. This choice of upper half plane is suggested by the general principle in field theory that m^2 should be treated as the limit of $(m^2 - i\varepsilon)$. In (4.29), this is equivalent to treating qE as limit $(qE + i\varepsilon)$, changing $\sin qas$ to $\sin(qa + i\varepsilon)s$. This makes the contour go above the poles. Equivalently, we can rotate the contour of integration in $L_{\rm eff}$ to the imaginary axis and express it in the alternative form:

$$L_{\rm eff} = -\int_0^\infty \frac{ds}{(4\pi)^2}\frac{\exp[-i(m^2 - i\varepsilon)s]}{s^3}\left(\frac{qas}{\sinh qas}\right)\left(\frac{qbs}{\sin qbs}\right). \tag{4.40}$$

This expression is sometimes easier to handle; it should be supplemented by the rule that poles along the real axis should be ignored by going below the axis.

The occurrence of the poles along the real axis and our $i\varepsilon$ – prescription has the following important consequence: It shows that $L_{\rm eff}$ has an imaginary part if a is non-zero. From (4.40) we get

$$\mathrm{Im}\ L_{\rm eff} = \int_0^\infty \frac{ds}{(4\pi)^2}\left(\frac{\sin m^2 s}{s^3}\right)\left(\frac{qas}{\sinh qas}\right)\left(\frac{qbs}{\sin qbs}\right) \tag{4.41}$$

and

$$\mathrm{Re}\ L_{\rm eff} = -\int_0^\infty \frac{ds}{(4\pi)^2}\left(\frac{\cos m^2 s}{s^3}\right)\left(\frac{qas}{\sinh qas}\right)\left(\frac{qbs}{\sin qbs}\right) \tag{4.42}$$

The expression in (4.41) can be evaluated by standard contour integration techniques. However, we can also calculate it from (4.39) directly; this calculation will explicitly show the origin of Im $L_{\rm eff}$. In (4.39),

$$L_{\rm eff}(E) = \int_0^\infty \frac{ds}{(4\pi)^2}\frac{\exp[-m^2 s]}{s^2}\left(\frac{qa}{\sin qas}\right)\left(\frac{qbs}{\sinh qbs}\right) \tag{4.43}$$

the poles at $s = s_n = (n\pi/qa)$ are to be avoided by going around small semicircles of radius ε in the upper half plane. The nth pole contributes to this semicircle the quantity

$$\begin{aligned} I_n &= \int_{\theta=\pi}^{\theta=0}\frac{(\varepsilon\exp(i\theta)i d\theta)}{(4\pi)^2 s_n^2}\exp[-m^2 s_n]\cdot\frac{qa}{\cos(n\pi)\cdot\varepsilon\exp(i\theta)}\left(\frac{qbs_n}{\sinh qbs_n}\right) \\ &= i(-1)^{n+1}\cdot\frac{(qa)^2}{16\pi^3}\left[\frac{1}{n^2}\exp\left(-\frac{m^2\pi}{qa}n\right)\right]\left(\frac{qbs_n}{\sinh qbs_n}\right). \end{aligned} \tag{4.44}$$

So the total contribution to Im $L_{\rm eff}$ is:

$$\mathrm{Im}\ L_{\rm eff} = \sum_{n=1}^{\infty}(-1)^{n+1}\cdot\frac{1}{2}\frac{(qa)^2}{(2\pi)^3}\frac{1}{n^2}\exp\left(-\frac{m^2\pi}{qa}n\right)\cdot\left(\frac{qbs_n}{\sinh qbs_n}\right). \tag{4.45}$$

It is now clear that Im $L_{\rm eff}$ arises because of non-zero a, i.e. whenever (i) there is an electric field in the direction of magnetic field or (ii) if **E** is perpendicular to **B**, but

$E^2 > B^2$. (In this case, we can go to frame in which the field is purely electric). For a purely electric field, the imaginary part is

$$\operatorname{Im} L_{\text{eff}} = \sum_{n=1}^{\infty} \frac{1}{2} \frac{(qE)^2}{(2\pi)^3} \frac{(-1)^{n+1}}{n^2} \exp\left(-\frac{\pi m^2}{qE} n \right). \tag{4.46}$$

Note that this expression is non-analytic in q; perturbation in powers of q will not produce this result.

4.2 *Effective lagrangian from path integral*

The above analysis relied heavily on the facts that: (i) the energy levels in a magnetic field are well known and (ii) the gauge and lorentz invariance of the theory puts severe restrictions of the form of L_{eff}. This method, therefore, is of only very limited validity. A more formal way of deriving this result will be to use the proper time representation for L_{eff} discussed in § 3·2. Since this gives a general formalism for handling arbitrary time dependence of the electric field, we will discuss this method next.

This method can produce both the Green's function and the effective lagrangian in a single stroke. The results quoted here will also be relevant for comparing the quantum theory in an arbitrary, time-dependent, electric field background with quantum theory in an expanding universe. The central quantity in this description is the kernel:

$$K(x, y; s) = \left\langle x \left| \exp\left[i\frac{s}{2}[(i\partial - qA)^2 - m^2 + i\varepsilon] \right] \right| y \right\rangle. \tag{4.47}$$

We saw in § 3.2 that the effective lagrangian L_{eff} and the propagator $G(x', x)$ can be calculated from this kernel by the relations

$$L_{\text{eff}} = -i \int_0^{\infty} \frac{ds}{s} K(x, x; s) \tag{4.48}$$

and

$$G(x', x) = \int_0^{\infty} ds\, K(x', x; s). \tag{4.49}$$

In the context of a scalar field interacting with an electromagnetic field, we can write the kernel in the form

$$K(x, y; s) = \langle x | \exp(ish) | y \rangle \tag{4.50}$$

with the 'Hamiltonian'

$$h = \tfrac{1}{2}(i\partial - qA)^2 - \frac{m^2}{2} + i\varepsilon. \tag{4.51}$$

We will consider an electric field along z-axis, which has an *arbitrary time dependence*; i.e. $\mathbf{E} = E(t)\hat{\mathbf{z}}, \mathbf{B} = 0$. The gauge is chosen such that $A^{\mu} = (0; 0, 0, A(t))$ [so that $A_{\nu} = (0; 0, 0, -A(t)); E(t) = -A'(t)$]. Using the translational invariance along the spatial coordinates, we can write

$$K(x^0, y^0; \mathbf{x}, \mathbf{x}; s) = \int \frac{d^3\mathbf{P}}{(2\pi)^3} \left\langle x^0 \middle| \exp\frac{is}{2}[(i\partial t)^2 - p_\perp^2 - (p_z - qA(t))^2 - m^2 + i\varepsilon] \middle| y^0 \right\rangle$$

$$= \int \frac{d^3\mathbf{p}}{(2\pi)^3} \exp\left(-\frac{is}{2}(p_\perp^2 + m^2 - i\varepsilon)\right)$$

$$\times \left\langle x^0 \middle| \exp\frac{is}{2}\left[-\frac{\partial^2}{\partial t^2} - (p_z - qA(t))^2 \right] \middle| y^0 \right\rangle$$

$$= \int \frac{d^3\mathbf{p}}{(2\pi)^3} \mathscr{G}(x^0, y^0; s) \exp[(-is/2)(p_\perp^2 + m^2 - i\varepsilon)] \tag{4.52}$$

where $\mathscr{G}(t, t'; s)$ is the propagator for the one-dimensional quantum mechanical problem with the hamiltonian

$$H = -\frac{1}{2}\frac{\partial^2}{\partial t^2} - \tfrac{1}{2}(p_z - qA(t))^2. \tag{4.53}$$

Let us now apply this formalism for the case of a uniform electric field for which the potential is $A = -Et$. Then

$$H = \frac{1}{2}\frac{\partial^2}{\partial t^2} - \tfrac{1}{2}(p_z + qEt)^2 = -\frac{1}{2}\frac{\partial^2}{\partial \rho^2} - \tfrac{1}{2}q^2E^2\rho^2 \tag{4.54}$$

where $\rho = t + (p_z/qE)$. This is a harmonic oscillator with mass $m = 1$ and *imaginary* frequency (iqE) ("inverted oscillator"). Since the path integral kernel for this problem is well-known we can immediately write down the coincidence limit for the propagator:

$$\mathscr{G}(t, t; s) = \left[\frac{qE}{(2\pi i \sinh qEs)}\right]^{1/2} \exp\left[-\frac{qE}{2i}\frac{2(\cosh qEs - 1)}{\sinh qEs}\left(t + \frac{p_z}{qE}\right)^2\right]$$

$$= \left[\frac{qE}{2\pi i \sinh qEs}\right]^{1/2} \exp\left[i\frac{qE}{\sinh qEs}(\cosh qEs - 1)\cdot\left(t + \frac{p_z}{qE}\right)^2\right]. \tag{4.55}$$

Doing the p_x, p_y and ω integrations, we are left with

$$K = \frac{qE}{(2\pi)^2}\cdot\left(\frac{1}{2is}\right)\cdot\frac{\exp((-is/2)(m^2 - i\varepsilon))}{\sinh(qEs/2)}$$

$$= \frac{1}{(2\pi)^2 is}\cdot\frac{(qE/s)}{\sinh(qEs/2)}\cdot\exp((-i/2)(m^2 - i\varepsilon)s). \tag{4.56}$$

Giving

$$L_{\text{eff}} = -i\int_0^\infty \frac{ds}{s}\cdot\frac{qE}{(2\pi)^2\cdot(2is)}\cdot\frac{\exp(-i(s/2))(m^2 - i\varepsilon)}{\sinh(qEs/2)}$$

$$= -\frac{1}{4}\int_0^\infty \frac{ds}{(2\pi)^2}\frac{1}{s^2}\frac{qE}{\sinh qEs}\exp(-i(m^2 - i\varepsilon)s)$$

$$= -\int_0^\infty \frac{ds}{4\pi^2}\cdot\frac{1}{s^2}\cdot\frac{qE}{\sinh qEs}\cdot\exp(-i(m^2 - i\varepsilon)s). \tag{4.57}$$

In this approach it is clear that the imaginary part arises because of the imaginary frequency (inverted nature) of the harmonic oscillator. This point is brought out more vividly by the corresponding calculation for the constant magnetic field. Magnetic fields, in general, give bounded hamiltonians. For example, consider the case with $A^\mu = (0; A(z), 0, 0)$ giving $B_y = (-\partial A/\partial z)$. Then,

$$K = \int \frac{d^2 p_\perp d\omega}{(2\pi)^3} \left\langle \exp\left(\frac{is}{2}[\omega^2 - p_y^2 - (p_x - qA)^2 + \partial_z^2 - m^2 + i\varepsilon]\right)\right\rangle$$

$$= \int \frac{d^2 p_\perp d\omega}{(2\pi)^3} \exp\left(+\frac{is}{2}(\omega^2 - p_y^2 - m^2 + i\varepsilon)\right)\mathscr{G}(z, z; s) \tag{4.58}$$

where the effective hamiltonian will be now

$$H = -\frac{1}{2}\frac{\partial^2}{\partial z^2} + \tfrac{1}{2}(p_x - qA(z))^2 \tag{4.59}$$

which has a potential bounded from below. Let us apply this equation to a uniform magnetic field; $A = -Bz$. Then

$$H = -\frac{1}{2}\frac{\partial^2}{\partial z^2} + \tfrac{1}{2}(p_x + qBz)^2 = -\frac{1}{2}\frac{\partial^2}{\partial \rho^2} + \tfrac{1}{2}q^2 B^2 \rho^2 \tag{4.60}$$

where $\rho = z + (p_x/aB)$. This is a harmonic oscillator with mass $m = 1$ and *real* frequency (qB). Therefore

$$\mathscr{G}(z, z; s) = \left[\frac{qB}{2\pi i \sin qBs}\right]^{1/2} \exp\left[i\frac{qB}{\sin qBs}(\cos qBs - 1)\cdot\left(z + \frac{p_x}{qB}\right)^2\right]. \tag{4.61}$$

Doing the p_x, p_y and ω integrations, we are left with

$$K = \frac{qB}{(2\pi)^2}\cdot\left(\frac{1}{2is}\right)\cdot\frac{\exp[-i(s/2)(m^2 - i\varepsilon)]}{\sin(qBs/2)}$$

$$= \frac{1}{(2\pi)^2 is}\cdot\frac{(qB/2)}{\sin(qBs/2)}\cdot\exp[-i(s/2)(m^2 - i\varepsilon)]. \tag{4.62}$$

Giving

$$L_{\text{eff}} = -i\int_0^\infty \frac{ds}{s}\cdot\frac{qB}{(2\pi)^2\cdot(2is)}\cdot\frac{\exp[-i(s/2)(m^2 - i\varepsilon)]}{\sin(qBs/2)}$$

$$= -\frac{1}{4}\int_0^\infty \frac{ds}{(2\pi)^2}\frac{1}{s^2}\frac{qB}{\sin qBs/2}\exp(-i(m^2 - i\varepsilon)s) \tag{4.63}$$

$$= -\int_0^\infty \frac{ds}{4\pi^2}\cdot\frac{1}{s^2}\cdot\frac{qB}{\sin qBs/2}\cdot\exp(-i(m^2 - i\varepsilon)s).$$

As it stands, the integrand has poles due to the $\sin(qBs/2)$ in the denominator. However, notice that the proper definition of harmonic oscillator path integral involves the prescription: $\omega = \lim_{\varepsilon\to 0}(\omega - i\varepsilon)$. Therefore in the kernel the factor $\sin qBs$ should be interpreted as the limit of the expression $\sin qBs(1 - i\varepsilon)$. So the poles are actually at

$$s_n = \pm\frac{2n\pi}{qB}(1 + i\varepsilon). \tag{4.64}$$

We can now transform the integral to one along the imaginary axis. Because of the $\exp[-is(m^2/2)]$ factor the contour should be closed in the lower half plane. Then we get

$$L_{\text{eff}} = -i\int_0^\infty \frac{ds}{s^2}K(s) = \int_0^\infty \frac{dy}{y^2}\frac{i\exp((-m^2/2)y)}{\sinh(qBy/2)}\cdot\frac{(qB/2)}{(2\pi)^2 i}$$

$$= \int_0^\infty \frac{dy}{y^2}\frac{\exp((-m^2/2)y)}{\sinh(qBy/2)}\cdot\frac{1}{(2\pi)^2}\cdot\left(\frac{qB}{2}\right). \qquad (4.65)$$

This expression, which is the same as (4.16), is real showing that the constant magnetic field does not create particles. We shall now discuss the renormalization of L_{eff}.

4.3 *Renormalization of the effective action*

We have seen earlier that the real and imaginary parts of the effective lagrangian lead to different classes of phenomena. Since the kernel is

$$K_{\text{total}} = \exp i\int [L_0(F) + L_{\text{eff}}(F)]\,d^4x$$

$$= \exp i\int [L_0(F) + \operatorname{Re} L_{\text{eff}}(F)]\exp(-\operatorname{Im} L_{\text{eff}})\,d^4x$$

$$= \langle 0, +\infty | 0, -\infty \rangle, \qquad (4.66)$$

we may interpret $\operatorname{Re} L_{\text{eff}}$ as a correction to the original lagrangian for the electromagnetic lagrangian

$$L_0(F) = \frac{1}{8\pi}(E^2 - B^2). \qquad (4.67)$$

The $(\operatorname{Im} L_{\text{eff}})$ is related to the probability for the system to make transitions from ground state to the excited state. In this particular case the excited state will be the one with the quanta of the scalar field present. We may, therefore, interpret, $2\operatorname{Im} L_{\text{eff}}$ as the probability per unit volume per unit time for production of scalar particles.

In this section we shall discuss the effects due to $\operatorname{Re} L_{\text{eff}}$. The pair creation probability arising from $\operatorname{Im} L_{\text{eff}}$ will be considered in the later sections.

The first point to note about $\operatorname{Re} L_{\text{eff}}$ is that it is divergent near $s = 0$. In fact, $\operatorname{Re} L_{\text{eff}}$ is divergent even when $\mathbf{E} = \mathbf{B} = 0$. This divergence—in accordance with the discussion we had before—must be spurious and can be removed by simply subtracting out the value for $\mathbf{E} = \mathbf{B} = 0$. Thus we modify (4.42) to

$$\operatorname{Re} L_{\text{eff}} \equiv R = -\int_0^\infty \frac{ds}{(4\pi)^2}\frac{\cos m^2 s}{s^3}\left[\frac{q^2 abs^2}{\sin qbs \sinh qas} - 1\right]. \qquad (4.68)$$

Since the subtracted term is a constant independent of $\mathbf{E}$, $\mathbf{B}$, the equations of motion are unaffected. The expression R is still logarithmically divergent near $s = 0$, since the quantity in the square brackets behaves as $[-\frac{1}{6}q^2 s^2(a^2 - b^2)]$ near $s = 0$. But notice that this divergent term is proportional to $(a^2 - b^2) = \mathbf{E}^2 - \mathbf{B}^2$, which is the original uncorrected, lagrangian. This opens up the possibility that we can reabsorb the divergence by redefining the field strengths, charges etc. This can be done as follows: Let us first write

$$L_{\text{total}} = L_0 + L_{\text{eff}} = (L_0 + L_c) + (L_{\text{eff}} - L_c) \tag{4.69}$$

where

$$L_c = -\frac{1}{(4\pi)^2}\int_0^\infty \frac{ds}{s^3}(\cos m^2 s)[-\tfrac{1}{6}(qs)^2(a^2-b^2)]$$

$$= \frac{q^2}{6(4\pi)^2}\int_0^\infty \frac{ds}{s}(\cos m^2 s)\cdot(a^2-b^2) \equiv \frac{Z}{8\pi}(a^2-b^2) = \frac{Z}{8\pi}(\mathbf{E}^2-\mathbf{B}^2) \tag{4.70}$$

with Z being a formally divergent quantity. With this trick, we can separate out the finite and divergent quantities in L_{total} and write

$$L_{\text{div}} = L_0 + L_c = \frac{1}{8\pi}(\mathbf{E}^2-\mathbf{B}^2) + \frac{Z}{8\pi}(\mathbf{E}^2-\mathbf{B}^2) = \frac{1}{8\pi}(1+Z)(\mathbf{E}^2-\mathbf{B}^2) \tag{4.71}$$

and

$$L_{\text{finite}} \equiv L_{\text{eff}} - L_c = -\frac{1}{(4\pi)^2}\int_0^\infty \frac{ds}{s^3}\cos m^2 s\left[\frac{q^2 s^2 ab}{\sin(qsb)\sinh(qsa)} - 1 + \frac{1}{6}q^2 s^2(a^2-b^2)\right]. \tag{4.72}$$

The quantity L_{finite} is perfectly well-defined and finite. [The leading term coming from the square bracket, near $s=0$ is proportional to s^3 and hence L_{finite} is finite near $s=0$.] So all the divergences are in the first term $(L_0+L_c)=(1+Z)L_0$. We shall now redefine all our field strengths and charges by the rule

$$\mathbf{E}_{\text{phy}} = (1+Z)^{1/2}\mathbf{E};\ \mathbf{B}_{\text{phy}} = (1+Z)^{1/2}\mathbf{B};\ q_{\text{phy}} = (1+Z)^{-1/2}q. \tag{4.73}$$

This is, of course, same as scaling a and b by $(1+Z)^{1/2}$ leaving $(q_{\text{phy}}E_{\text{phy}}) = qE$ invariant. Since only the products qa, qb appear in L_{finite}, it can also be expressed in terms of $(q_{\text{phy}}E_{\text{phy}})$. Thus it is possible to redefine the variables in our theory, thereby absorbing the divergent quantities. The remaining expressing L_{finite} is well-defined and possesses a Taylor expansion in q^2. Using this expansion, one can calculate corrections to the electromagnetic lagrangian in an order-by-order manner (see Heisenberg and Euler 1936; Schwinger 1954a,b).

4.4 *Quantization in a time-dependent gauge: Bogoliubov coefficients*

In the discussion so far, we have derived the form of the effective action and studied its real part. We shall now consider the physical origin of the imaginary part to L_{eff}. The existence of an imaginary part to L_{eff} suggests that the probability for the quantum system [here, the scalar field] to be in the ground state at $t=\infty$ is less than unity. As the excited states of the scalar field can be interpreted as states containing non-zero number of scalar quanta, this phenomenon can be thought of as particle creation by the electric field. Since the notion of a *static* electromagnetic field creating particles may be rather surprising, we will examine the origin of this phenomenon more closely (Schwinger 1954a,b; Nikishov 1970a,b; Popov 1972; also see articles in Ginzburg 1987). As we shall see, there are some interesting conceptual issues, connected with gauge invariance of this phenomenon, which needs to be scrutinized carefully (Padmanabhan 1990, 1991a). To do this, we will describe a constant electric field in two different gauges, one in which the potential is time dependent and the other in which it is not. The quantum theory of the scalar field will be studied in these two gauges in §§4.4 and 4.5 and the results will be compared in §4.6.

Let us begin by quantizing a complex scalar field ϕ in the gauge $A_i^{(1)}$: In this (time-dependent) gauge, the constant electric field $\mathbf{E}=(E,0,0)$ is represented by the vector potential $A_\mu=(0,Et,0,0)$; with $A^\mu=(\phi,\mathbf{A})=(0,-Et,0,0)$. Since the Klein-Gordon equation

$$[(i\partial_k - qA_k)^2 - m^2]\phi(t,\mathbf{x}) = 0 \tag{4.74}$$

has an explicit time dependence in this gauge it is not easy to provide a particle interpretation. The usual strategy adopted in such cases is the following (Parker 1982): (i) We obtain a complete set of orthonormal solutions to (4.74) which can be identified as positive and negative frequency modes in the asymptotic past, i.e. as $t\to-\infty$. This task itself is somewhat tricky since the field does not vanish asymptotically; we will have to identify as positive frequency modes those solutions which have decreasing phase in the adiabatic limit. (This is equivalent to choosing the modes as those will behave as $\exp(-i\omega t)$ in the limit of vanishing E.) (ii) We can also obtain, in a similar manner, the positive and negative frequency modes for the asymptotic future. Because of the time dependence of A_i, a mode which is purely positive frequency in the infinite past will evolve into a combination of positive and negative frequencies in the infinite future, a phenomenon which is usually interpreted as pair creation. [This prescription is not as well defined as one would like it to be; see Parker 1982].

The mode functions for the Klein-Gordon equation can then be expressed in the form:

$$\phi(t,\mathbf{x}) = f_{\mathbf{k}}(t)\exp(i\mathbf{k}\cdot x) \tag{4.75}$$

where $f_k(t)$ satisfies the equation

$$\frac{d^2}{dt^2}f_k + [m^2 + k_\perp^2 + (k_x + qEt)^2]f_k = 0; \quad \mathbf{k}_\perp = (k_y, k_z). \tag{4.76}$$

Introducing the variables,

$$\tau = (qE)^{1/2}t + \frac{k_x}{(qE)^{1/2}}; \; \lambda = \frac{1}{qE}(m^2 + k_\perp^2); \; \nu = -\tfrac{1}{2}(1 - i\lambda) \tag{4.77}$$

this equation becomes

$$\ddot{f}_k + (\tau^2 + \lambda)f_k = 0 \tag{4.78}$$

which is essentially a Schrödinger equation in an inverted oscillator potential. If $f(\lambda,\tau)$ is a solution, then so are the functions $f^*(\lambda;\tau)$, $f(\lambda;-\tau)$ and $f(\lambda;-\tau)^*$. This solution set can be taken to be

$$\{D_{\nu^*}((1+i)\tau),\; D_\nu(1-i)\tau),\; D_{\nu^*}(-(1+i)\tau),\; D_\nu(-(1-i)\tau)\} \tag{4.79}$$

where $D_\nu(z)$ is the parabolic cylinder function. Only two of these four functions are linearly independent.

To proceed further we have to identify the 'positive' frequency modes in the asymptotic past. Since the equation does not admit $\exp(\pm i\omega t)$ type solutions in the asymptotic limit, we have to use some other criterion. The usual approach is to take the WKB solution and identify the positive and negative modes by looking at the phase of the solutions. In the WKB region, the solutions are

$$f(\tau) \approx \frac{1}{[p(\tau)]^{1/2}} \exp\left(\pm i \int p d\tau \right); \; p = (\tau^2 + \lambda)^{1/2}. \tag{4.80}$$

This gives

$$f(\tau) \approx |\tau|^{\pm i\lambda/2 - 1/2} \exp(\pm (i/2)\tau^2) \tag{4.81}$$

so that the two independent solutions can be taken to be

$$f_+(\tau) = |\tau|^{\nu} \exp((i/2)\tau^2); \; f_-(\tau) = |\tau|^{\nu^*} \exp(-(i/2)\tau^2) = [f_+(\tau)]^*. \tag{4.82}$$

The solution with decreasing phase is $f_-(\tau)$ for $\tau > 0$ and $f_+(\tau)$ for $\tau < 0$. [These are, of course, limiting forms of $D_\nu(z)$. For $\tau \to -\infty$, the positive and negative frequency solutions are $D_\nu[-(1-i)\tau]$ and $D_{\nu^*}[-(1+i)\tau]$ while for $\tau \to +\infty$, the positive and negative frequency modes are $D_{\nu^*}[(1+i)\tau]$ and $D_\nu[1-i)\tau]]$.

Let us consider the evolution of these modes. If we start with a pure f_+ mode in the distant past, it will evolve into a linear combination of positive and negative frequency modes in the distant future. So we may write, in general, a solution of the form:

$$\psi(\tau) = \begin{cases} |\tau|^{\nu} \exp(i/2)\tau^2 & \tau \to -\infty \\ A\tau^{\nu} \exp(i/2)\tau^2 + B\tau^{\nu^*} \exp(-(i/2)\tau^2) & \tau \to +\infty \end{cases}. \tag{4.83}$$

This is equivalent to a scattering problem in an inverted oscillator potential $V(\tau) = (-\frac{1}{2}\omega^2\tau^2)$ in which a wave with amplitude B is incident from right, and is transmitted with unit amplitude and reflected with an amplitude A. The quantity $|A|^2$ determines the overlap between positive and negative frequency modes. We can compute this quantity using another trick due to Landau (Landau and Lifshitz 1973). Treating τ as the real part of a complex number, one can see that rotating τ in complex plane from $\theta = 0$ to $\theta = \pi$, maps the $\exp((i/2)\tau^2)$ part of the solutions into each other. This immediately gives

$$\exp(i\pi\nu) = A = \exp[i\pi(-\tfrac{1}{2}(1 - i\lambda))] = \exp(-i\pi/2)\exp(-\pi\lambda/2) \tag{4.84}$$

so that $|A|^2 = \exp(-\pi\lambda)$. Thus a positive frequency mode at $\tau \ll 0$ with an amplitude unity picks up an amplitude A to have negative frequency for $\tau \gg 0$. Hence the cross term in the Bogoliubov coefficients is just A. So,

$$\{\text{mean number of particles created in mode } \lambda\} = \bar{n}_\lambda = |A|^2 = \exp(-\pi\lambda). \tag{4.85}$$

The normalization condition, $|B|^2 - |A|^2 = 1$, for Bogoliubov coefficients (which, in the present context, is related to the conservation of probability for an equivalent Schrödinger equation) gives

$$|B|^2 = 1 + |A|^2 = 1 + \exp(-\pi\lambda). \tag{4.86}$$

[A more formal way of deriving A and B is discussed in Appendix 1]. The relative probability for pair creation in mode-λ is

$$R_\lambda = \frac{|A|^2}{|B|^2} = \frac{\exp(-\pi\lambda)}{1+\exp(-\pi\lambda)} = (\exp(\pi\lambda)+1)^{-1}.$$

The probability that *no* pair creation occurs is

$$P_\lambda = 1 - R_\lambda = \frac{\exp(\pi\lambda)}{1+\exp(\pi\lambda)} = \frac{1}{1+\exp(-\pi\lambda)} = \frac{1}{1+\bar{n}_\lambda}. \tag{4.87}$$

Therefore the 'vacuum persistence probability' will be

$$|\langle \text{out}, 0|0, \text{in}\rangle|^2 = \prod_\lambda P_\lambda = \prod_\lambda \frac{1}{(1+\bar{n}_\lambda)} = \exp\left[-\sum_\lambda \ln(1+\bar{n}_\lambda)\right]$$
$$\equiv \exp\left[-2\int d^4x \operatorname{Im} \mathscr{L}_{\text{eff}}\right] \tag{4.88}$$

where, in the last line, we have introduced the imaginary part of the electromagnetic effective action in the standard manner. This allows us to identify

$$2\int d^4x \operatorname{Im}\mathscr{L}_{\text{eff}} = \sum_\lambda \ln(1+\exp(-\pi\lambda)) = \sum_{\lambda, N} (-1)^{N+1}\frac{1}{N}\exp(-\pi N\lambda). \tag{4.89}$$

Changing the summation to an integration by the rule

$$\sum_\lambda \to V\int \frac{dk_y}{2\pi}\frac{dk_z}{2\pi}\frac{dk_x}{2\pi} = \frac{V}{(2\pi)^3}\int dk_x \cdot \int_0^\infty 2\pi k_\perp dk_\perp \tag{4.90}$$

we can rewrite the N-th term as

$$\frac{(-1)^{N+1}}{N}\frac{V}{(2\pi)^3}\cdot\int dk_x \cdot \int_0^\infty \pi d(k_\perp^2)\exp\left[-\frac{\pi N}{qE}(m^2+k_\perp^2)\right]$$
$$= \frac{(-1)^{N+1}}{N}\frac{V}{(2\pi)^3}\int dk_x . \pi . \left(\frac{qE}{\pi N}\right)\exp\left(-\frac{\pi N}{qE}m^2\right)$$
$$= \frac{(qE)V}{(2\pi)^3}\int dk_x \frac{(-1)^{N+1}}{N^2}\exp\left(-\frac{\pi m^2}{qE}N\right)$$
$$= \frac{(qE)^2 VT}{(2\pi)^3}\frac{(-1)^{N+1}}{N^2}\exp\left(-\frac{\pi m^2}{qE}N\right). \tag{4.91}$$

In arriving at the last expression, we have interpreted a $\delta(0)$ as giving rate per unit volume per unit time of physical process; since k_x and $(qE)t$ have the same dimensions the last integral is performed over $(qE)t$ for some finite time interval. We thus get the final result

$$\operatorname{Im} L_{\text{eff}} = \sum_{n=1}^\infty \frac{1}{2}\frac{(qE)^2}{(2\pi)^3}\frac{(-1)^{n+1}}{n^2}\exp\left(-\frac{\pi m^2}{qE}n\right) \tag{4.92}$$

which is the same as the one obtained before. Some of the manipulations leading to this result can be made more precise by introducing an adiabatic "switch-off" of the electric field. But the final results will be the same

4.5 *Quantization in space dependent gauge: Tunnelling*

Let us now consider the same physical system in the space-dependent gauge. A difficulty arises when we repeat the same analysis in the space-dependent gauge with $A_\mu = (-Ex, 0, 0, 0) = A^\nu$. Since the vector potential is now independent of time, it is obvious that the solutions to (4.74) can be expressed in the form

$$\phi(t, \mathbf{x}) = \sum_{\mathbf{k},\omega} (A_{\mathbf{k}} g_{\mathbf{k},\omega}(x) \exp(-i\omega t + ik_y y + ik_z z) + B_{\mathbf{k}}^{\dagger} g^{*}_{\mathbf{k},\omega}(x) \exp(i\omega t - ik_y y - ik_z z) \tag{4.93}$$

where $\mathbf{k} = (k_y, k_z)$ and the function g satisfies the equation

$$\frac{d^2 g_{\mathbf{k},\omega}}{dx^2} + [(\omega + qEx)^2 - |\mathbf{k}|^2 - m^2] g_{\mathbf{k},\omega} = 0 \tag{4.94}$$

which can be again solved in terms of parabolic cylinder functions. The key point to note is that the time dependence of the mode functions in this gauge is just $\exp \pm i\omega t$ at all times. If we use $A_{\mathbf{k},\omega}$ (and its hermitian conjugate) as the annihilation (and creation) operator for our particles, such particles are not produced by the constant electric field. The vacuum state defined by these modes remains as vacuum for all times.

Let us look at this situation more closely. Substituting

$$\rho = (qE)^{1/2} x + \frac{\omega}{(qE)^{1/2}};\ \lambda = \frac{1}{qE}(m^2 + k_\perp^2);\ \nu = -\tfrac{1}{2}(1 - i\lambda). \tag{4.95}$$

The equation for g becomes:

$$\frac{d^2 g}{d\rho^2} + (\rho^2 - \lambda) g = 0. \tag{4.96}$$

This is the same equation we met before with the sign of λ changed; this change is equivalent to $\nu \leftrightarrow \nu^*$. So the solution set is still the same with some change of signs:

$$\{D_\nu((1+i)\rho), D_{\nu^*}((1-i)\rho), D_\nu(-(1+i)\rho), D_{\nu^*}(-(1-i)\rho)\}. \tag{4.97}$$

The first pair can be interpreted as the negative and positive frequency modes in the far right ($\rho \to +\infty$) while the second pair corresponds to positive and negative frequency modes in the far left. [We define positive frequency as the one with decreasing phase *in time* but increasing phase in space. So the negative frequency in time will become positive "frequency" in space with the $\nu \leftrightarrow \nu^*$ change]. A meaningful theory can be constructed out of any independent pair of these solutions. We will not obtain any pair creation in the manner in which we obtained it earlier.

What is usually done in the literature at this stage is the following: Since the natural definition of particles in the far left does not match with the natural definition of particles in the far right, one can attempt an interpretation for particle creation in terms of 'tunnelling' across the potential. [To be precise, what we will be concerned with is not tunnelling but its close relative, 'over-the-barrier- reflection'; see Marinov and Popov 1977; Landau and Lifshitz 1973 (p. 190)]. This approach leads to the same result as before. To see this, consider a mode which is right-moving in $\rho > 0$ region. [i.e., positive frequency for $\rho \to +\infty$]. This is given by $D_{\nu^*}((1-i)\rho)$. We look at its behaviour in the left, $\rho \to -\infty$; we can express it as a superposition of $D_\nu(-(1+i)\rho)$ and $D_{\nu^*}(-(1-i)\rho)$. Using the relation

$$D_p(z) = \exp(i\pi p)\, D_p(-z) + \frac{(2\pi)^{1/2}}{\Gamma(-p)} \exp\left(\frac{i\pi}{2}(p+1)\right) D_{-p-1}(-iz) \tag{4.98}$$

we get

$$D_{\nu^*}((1-i)\rho) = \exp(i\pi\nu^*)\, D_{\nu^*}(-(1-i)\rho) + \frac{(2\pi)^{1/2}}{\Gamma(-\nu^*)} \times \exp\left(\frac{i\pi}{2}(\nu^*+1)\right) D_\nu(-(1+i)\rho). \tag{4.99}$$

Asymptotically as $\rho \to \infty$

$$\begin{aligned}\psi = D_{\nu^*}((1-i)\rho) &\cong (\sqrt{2})^{\nu^*} \rho^{\nu^*} \exp\left(\frac{i}{2}\rho^2\right) \exp\left(-\frac{i\pi}{4}\nu^*\right)\\ &\equiv B\rho^{\nu^*} \exp\left(\frac{i}{2}\rho^2\right)\end{aligned} \tag{4.100}$$

while, as $\rho \to -\infty$,

$$\begin{aligned}\psi &= \exp(i\pi\nu^*)\left\{(2\pi)^{\nu^*/2} |\rho|^{\nu^*} \exp\left(-\frac{i\pi}{4}\nu^*\right) \exp\left(\frac{i}{2}\rho^2\right)\right\}\\ &\quad + \frac{(2\pi)^{1/2}}{\Gamma(-\nu^*)} \exp\left(\frac{i\pi}{2}(\nu^*+1)\right) \left\{(\sqrt{2})^{\nu} |\rho|^{\nu} \exp\left(\frac{i\pi}{4}\nu\right) \exp\left(\frac{i}{2}\rho^2\right)\right\}\\ &\equiv A|\rho|^{\nu^*} \exp\left(\frac{i}{2}\rho^2\right) + C|\rho|^{\nu} \exp\left(-\frac{i}{2}\rho^2\right).\end{aligned} \tag{4.101}$$

We can identify the transmission and reflection coefficients as

$$T = \frac{B}{A};\; R = \frac{C}{A}. \tag{4.102}$$

A simple calculation now gives

$$|T|^2 = \exp(-\pi\lambda);\, |R|^2 = 1 + \exp(-\pi\lambda). \tag{4.103}$$

We note that the reflection coefficient is greater than unity signalling the well-known Klein's paradox (Fulling 1989). The transmission factor and the excess over unity of the reflection factor are attributed to the pairs created by the field. These expressions clearly agree with the results obtained in the time-dependent gauge. [This idea can be cast into a more formal language; the key point is that when A_0 is non-zero, positive frequency modes and positive norm modes need not be the same; for a detailed discussion see Fulling (1989)].

This is the conventional interpretation. We will examine this situation more closely in the next section; before that, it is worthwhile to see what happens to the path integral kernel in this gauge. If we take $\mathbf{E}=E(z)\hat{z}$, $\mathbf{B}=0$ in the gauge $A^\mu=(A(z),0,0,0)$ the path integral kernel becomes

$$\begin{aligned}
K(x^0,x_\perp,z;x^0,x_\perp,z';s) &= \\
&= \int\frac{d^2p_\perp}{(2\pi)^2}\frac{d\omega}{(2\pi)}\left\langle x^i\left|\exp\left\{\frac{is}{2}[(\omega-qA)^2-p_\perp^2+\partial_z^2-m^2+i\varepsilon]\right\}\right|y^i\right\rangle \\
&= \int\frac{d^2p_\perp d\omega}{(2\pi)^3}\exp\left[\frac{-is}{2}(p_\perp^2+m^2-i\varepsilon)\right]\left\langle z\left|\exp\left\{\frac{is}{2}[\partial_z^2+(\omega-qA)^2]\right\}\right|z'\right\rangle \\
&= \int\frac{d^2p_\perp d\omega}{(2\pi)^3}\exp\left[-\frac{is}{2}(p_\perp^2+m^2-i\varepsilon)\right]\mathscr{G}(z,z;s)
\end{aligned} \tag{4.104}$$

with the hamiltonian

$$H=-\frac{1}{2}\frac{\partial^2}{\partial z^2}-\tfrac{1}{2}(\omega-qA(z))^2. \tag{4.105}$$

This is the same hamiltonian as before with the replacement in the dummy variable: $t\Leftrightarrow z$. It is now obvious that L_{eff} calculated here will be the same as the one computed in the time-dependent gauge as long as the field is constant in space and time.

Incidentally, this analysis also reveals another interesting fact. The L_{eff} will always be the same for the two vector potentials: $A^i=(f(z),0,0,0)$ and $A^i=(0,0,0,f(t))$ where f in both cases is the same function of its argument. When f is a linear function, both these gauges correspond to a constant, homogeneous electric field. *But notice that, in general, the electric fields arising from these two potentials are quite different. Nevertheless the particle production in these two fields will be the same.* This curious fact does not seem to have been noticed in the literature.

4.6 *Comparison of the two gauges*

The study in the previous sections reveals the essential differences and similarities in the quantization of the scalar field in two different gauges. We will now compare these results (Padmanabhan 1990, 1991a).

It is important to realize that the "particles" (and, of course, the vacuum state) defined in these two gauges are quite different. To begin with, it must be clear that the *mode functions in these two gauges do not refer to the same 'particles'*. This can be seen in the following manner: The positive frequency mode in the $A_i^{(2)}$ gauge

$$\phi^{+,\text{trans}}_{\mathbf{k},\omega}(t,\mathbf{x})=F_{\mathbf{k},\omega}(x)\exp(-i\omega t+ik_y y+ik_z z) \tag{4.106}$$

can be gauge transformed into the $A_i^{(1)}$ gauge by the usual addition of the phase. We will then get,

$$\phi^{+,\text{trans}}_{\mathbf{k},\omega}(t,\mathbf{x})=F_{\mathbf{k},\omega}(x)\exp(-i(\omega+qEx)t+ik_y y+ik_z z) \tag{4.107}$$

which is not a pure positive frequency mode in the $A_i^{(1)}$ gauge.

Since the choice of positive frequency modes in these two gauges do not match, the vacuum state defined in one gauge will contain the 'particles' defined by the modes in the other gauge. This particle content can be ascertained by evaluating the projection of the (gauge transformed) positive frequency modes $\phi_{\mathbf{k},\omega}^{+,\mathrm{trans}}(t, \mathbf{x})$ onto the negative frequency modes $\psi_{\mathbf{k}}^{-}(t, \mathbf{x}) \equiv g_k^{-}(t) \exp i\mathbf{k}\cdot\mathbf{x}$ or $\eta_{\mathbf{k}}^{-}(t, \mathbf{k}) \equiv f_k^{-}(t) \exp i\mathbf{k}\cdot\mathbf{x}$, using the conserved, gauge invariant Hilbert space scalar product

$$(f, g) = i \int \mathrm{d}\mathbf{x} [f^*(\partial_0 + iqA_0)g - g(\partial_0 - iqA_0)f^*]. \tag{4.108}$$

Since the negative frequency mode of infinite past η^- is not the same as that in the infinite future ψ^-, it is clear that we cannot have *both* the products $(\eta^-, \phi_{\mathrm{trans}}^+)$ and $(\psi^-, \phi_{\mathrm{trans}}^+)$ vanishing. It is, for example, possible to choose the solution to (4.94) in such a way that $(\eta^-, \phi_{\mathrm{trans}}^+) = 0$; and $(\eta^+, \phi_{\mathrm{trans}}^+) = 1$. This ensures that the concept of the particle (and vacuum state) defined in the static gauge $A_i^{(2)}$ is the same as the concept of particle defined in the infinite past in the gauge $A_i^{(1)}$. A complicated but straightforward calculation will then show that such a choice leads to the result

$$|(\psi^-, \phi_{\mathrm{trans}}^+)|^2 = |c_2|^2 = n_{\mathbf{k}} \tag{4.109}$$

in the infinite future.

This result may be summarized as follows: In the static gauge, one may define the particle concept in a time independent manner; vacuum state remains a vacuum. In the gauge $A_i^{(1)}$ it is not possible to provide a particle definition except asymptotically. We can, at best, make these definitions match at one instant, say, in the asymtotic past. Then, in the asymptotic future, the vacuum state will remain a vacuum state as far as particles of gauge $A_i^{(2)}$ are concerned but will be populated by the particles of gauge $A_i^{(1)}$.

The inequivalence can also be seen by comparing the Green's functions (constructed out of our choice of mode functions). Let $G_1(x, x')$ be the Green's function in the time dependent gauge. Under the gauge transformation with a gauge function $f(x)$, the Green's function gets multiplied by the factor $\exp iq(f(x) - f(x'))$. The conventional proofs of the gauge invariance of quantum electrodynamics is based on the tacit assumption that the Green's functions are transformed in the above manner. But in our case the Green's function $G_2(x, x')$ in the space dependent gauge is *not* the one obtained by gauge transforming $G_1(x, x')$. [This is trivial to see by comparing, say, the time dependence of the two Green's functions. Of course, the result is obvious from the fact that the mode functions in the space dependent gauge are not the ones obtained from the mode functions in the time dependent gauge by the standard gauge transformation. Since both the Green's functions satisfy the same equations they essentially differ by the choice of boundary conditions.] It is well known from standard analysis in field theory that the Green's functions carry complete information about the vacuum state and particle spectrum of the theory. Thus we can conclude that the particles defined in these two gauges are not physically the same.

It is also possible to construct model "particle detectors" to provide an alternative definition of particles. These detectors essentially measure the temporal Fourier transform of the Green's functions. Such an analysis leads to the same conclusions as above.

One may now ask: What about the calculation of L_{eff} in the two gauges? Why do these calculations give the same result? The reason for this result has to do with the tacit boundary conditions assumed in Schwinger's proper time approach. The entire philosophy behind the computation of L_{eff} from the kernel is based on the assumption that the fields are switched off in the asymptotic past and future. This fact is not of any special significance in the time dependent gauge because, in that gauge, we were actually handling an electric field which was varying with time in an *arbitrary* manner. While repeating the calculation in the space dependent gauge (using Schwinger's approach) we have tacitly assumed that this field also has a hidden time dependence which makes it vanish in the asymptotic limits. Thus we have tacitly assumed that the 'in' and 'out' vacua of the space dependent gauge are defined at late times when the field is switched off. These vacua are the same as the ones in the time dependent gauge with the fields switched off asymptotically. It is, therefore, not surprising that we get the same particle creation rates in both the gauges if we use the path integral approach. Schwinger's method is ingenious in the sense that it automatically takes into account this boundary condition. But in our discussion in §§4.4 and 4.5 we explicitly constructed the mode functions without assuming that the field vanishes asymptotically. Quite clearly, the concept of particles in this case is different from the one obtained by assuming that the fields vanish asymptotically.

Notice that, even in the path integral approach, it is only the coincidence limit of the kernel, $K(x,x;s)$ which is the same in both the gauges. The kernel for arbitrary points $K(x,y;s)$ computed in the two gauges are inequivalent—in the sense that they are *not* related to each other by a gauge transformation. It is this feature which makes, for example, the Green's functions in the two gauges inequivalent.

It is worthwhile to ask: what happened to the gauge invariance of the theory in the case of a constant electric field which is *not* switched off asymptotically? In the classical theory, we are allowed to make gauge transformations $A_i \rightarrow A_i + \partial_i \chi$ with any sufficiently smooth χ. However, this is not the case in the quantum theory. The allowed class of gauge transformations are now only those which can be implemented as unitary transformations in the Hilbert space of states. This will necessarily impose some constraints on the global, asymptotic behaviour of the allowed set of functions χ which induces gauge transformations. It can be shown that the gauge transformation we are considering does not belong to this set. (For a detailed discussion of this issue, see Nenciu and Scharf 1978; Seip 1982 and the references cited therein. For a possible approach to construct theories which depends only on the electromagnetic fields and not on A_i, in this specific context, see Capri and Roy 1991). Thus the theories based on gauges $A_i^{(1)}$ and $A_i^{(2)}$ live in different Hilbert spaces; there is no unitary transformation connecting them. [Classically too, there can be situations in which a formal transformation is not canonically implementable. Possibly there is an interesting avenue here which needs to be explored.]

The overall picture which emerges from the above analysis is the following: If the electric and magnetic fields vanish asymptotically in space and time, then particles can be defined unambiguously in these asymptotic regions. (In these regions, we set $A_\mu = 0$.) The S-matrix elements describing the transition between the asymptotic states is well defined and gauge invariant. This is probably the only situation in which an unambiguous statement can be made. If the field does not vanish asymptotically or if we desire to have a particle definition in the "strong field region" (rather than be contend with the particles defined asymptotically), then we have to necessarily specify the vector potential as well. It is possible that the particle definition remains the same for a *subclass* of vector potentials connected by gauge transformations. Whether such a subclass can be specified in any meaningful way is not known (at least to the author) and seems worth investigating. It is also possible to construct toy models which serve as detectors for particles and use them to define the concept of particle. In general, such a definition does not agree with the definition based on quantum states. However, it is probably worth studying the response of detectors and compare them with the results obtained above. Note that if the detector is made of normal physical systems and coupled to electromagnetic field in a gauge invariant way then—by very definition—it will give a gauge invariant response. The ambiguity will not be in the output we receive from the detector but in translating this output to the concept of particles. (There is also some issues of principle involved in describing operationally the concept of an electromagnetic field *in a particular* gauge; we will comment on this issue at the end of § 4.7) These questions are under study.

4.7 *Quantum theory in a singular gauge*

We shall now consider quantum theory in a very different kind of background (Padmanabhan 1991a). These backgrounds, which involve description of nonsingular fields in singular gauges, occur both in electromagnetism and gravity. We shall consider the electromagnetic case here; the analogous situation in gravity will be considered in § 5.3.

Since we expect physical processes (like pair creation) to be invariant under the gauge transformation, $A_\mu \to A_\mu + \partial_\mu f$, (at least for a class of functions) we expect L_{eff} to depend only on $F_{\mu\nu}$ and not on the gauge chosen to describe the field. It is interesting to see how this result comes about in the path integral approach. Under the gauge transformation $A_\mu \to A_\mu + \partial_\mu f$, the term $(qA_\mu \dot{x}^\mu)$ in the lagrangian picks up the additional piece $(q\dot{x}^\mu \partial_\mu f)$ and the path integral amplitude is multiplied by the factor

$$P = \exp iq \int_0^s \mathrm{d}\tau (\dot{x}^\mu \partial_\mu f). \tag{4.110}$$

It is usual to 'perform' the integration in the above expression and obtain

$$P = \exp iq \int_0^s \mathrm{d}\tau \frac{\mathrm{d}}{\mathrm{d}\tau} f[x^\mu(\tau)] = \exp iq[f(x') - f(x)]. \tag{4.111}$$

If this result is valid, then the physics will be gauge invariant. The L_{eff} only involves the coincidence limit $K(x, x; s)$ for which–assuming we can take $f(x') = f(x)$ when $x = x'$—the factor P is unity. The propagator is only modified by a phase, and it can be shown that amplitudes for physical processes do not change.

The above discussion can be cast in a different form which is more useful in what follows. We may say that path integral amplitude for a given path $x^i(\tau)$, connecting x and x' gets multiplied by a factor

$$F = \exp iq \int_x^{x'} \mathrm{d}\tau \dot{x}^\mu A_\mu \tag{4.112}$$

in the presence of the electromagnetic field. If we change the gauge, F will be further multiplied by P in (4.110). As long as we can integrate (4.110) to obtain (4.111), each amplitude will be multiplied by a factor which is independent of the path (and dependent only on the end points); so the physics will remain unchanged.

There are, however, situations in which the result (4.111) cannot be obtained from (4.110). It is precisely these cases which provide interesting analogies with phenomena in curved spacetime and accelerated frames.

The derivation leading to (4.111) fails in two physically interesting cases. The first one—which is well-known in a different grab—is the following (Padmanabhan 1991b): Suppose one of the spatial coordinates we are using is periodic but $f(x)$ does not respect this periodicity. For example, we may use the $x^i = (t, r, \theta, z)$ coordinate system and take $f(x^i) = B\theta$ [clearly, $f(\theta + 2\pi n) \neq f(\theta)$]. A path which "winds" around the origin in xy-plane n-times will now produce an additional factor

$$F_n \equiv \exp iqB(\theta - \theta' + 2\pi n). \tag{4.113}$$

The original kernel could have been written as

$$K_{B=0}(x', x; s) = \sum_{n=-\infty}^{\infty} K_n(x', x; s) \tag{4.114}$$

where K_n is the kernel obtained by summing over paths with a given 'winding number' n. The kernel in the presence of a gauge-function f will be

$$K_B(x', x; s) = \sum_{n=-\infty}^{\infty} K_n(x', x; s) \exp[iqB(\theta - \theta' + 2\pi n)]. \tag{4.115}$$

This will modify the L_{eff} as well, since the coincidence limit is also affected.

$$K_B(x', x; s) = \sum_{n=-\infty}^{\infty} K_n(x, x; s) \exp[-iqB2\pi n)]. \tag{4.116}$$

Notice that the kernel invariant under the change $\theta' \to \theta' + (2\pi/qB)$. Thus θ now has a periodicity of $(2\pi/qB)$. It is clear that the propagator $G(x', x)$ will also exhibit this periodicity.

This above example is essentially a calculation of K (and L_{eff}) for an Aharonov-Bohm potential. It is clear that, even though A_μ appears to be expressible as $(\partial_\mu f)$ locally, we do not have a pure-gauge situation. The ill-defined nature of f at origin leads to a delta-function magnetic field along z-axis (Aharonov-Bohm field). This is most easily seen by noticing that the flux through a path around the origin

$$\oint A_\mu dx^\mu = \int_0^{2\pi} A_\theta d\theta = 2\pi B \tag{4.117}$$

is non-zero. Paths with different winding numbers cannot be continuously deformed to each other. Note that the f and A_μ in this case can be written in the Cartesian coordiantes as

$$f = B \tan^{-1}\left(\frac{y}{x}\right); A_\mu = B\left(0, \frac{y}{x^2 + y^2}, +\frac{x}{x^2 + y^2}, 0\right). \tag{4.118}$$

Let us now look at a more unfamiliar situation (which is of greater importance in what follows) in which (4.111) fails. Consider a different vector potential:

$$A_\mu = B\left(0, \frac{y}{x^2 - y^2}, +\frac{x}{x^2 - y^2}, 0\right) \tag{4.119}$$

which can be obtained from the gauge function

$$f(x) = \begin{cases} B\tanh^{-1}(y/x); & |y| \leqslant |x| \\ B\tanh^{-1}(x/y); & |y| \geqslant |x| \end{cases}. \qquad (4.120)$$

In the previous case, $f(x)$ was multivalued; in the present case, it is *singular* at $x = \pm y$. The function A_μ is badly divergent on the planes $x = \pm y$. The integral I which determines the phase factor for the path,

$$I = \int_0^s \mathrm{d}\tau \dot{x}^\mu A_\mu[x^\mu(\tau)] \qquad (4.121)$$

will give a well-defined answer $[f(x(s)) - f(x(0))]$ *only if* the path $x^\mu(s)$ does *not* cross the singular planes $x = \pm y$. But in the path integral kernel, we have to sum over *all* paths whatever may be the end points. In such a sum all paths which cross this surface will give a divergent contribution, thereby making the entire kernel ill-defined.

We thus reach the conclusion that if the gauge function diverges on some surface which divides the spacetime into two regions, then the kernel cannot be defined. In fact we cannot even define the amplitude for any path which crosses the singular surface; the integral in (4.112) cannot be evaluated across the singularity. This is, of course, a more serious situation than when gauge function becomes ill-defined only on a point; in that case we could manage to define the kernel by including the appropriate phase.

One important consequence of this result is the following: We know that the semiclassical propagator can be represented as

$$\mathscr{G}(x, x') = N \exp iS(x, x') \qquad (4.122)$$

where S is the classical action satisfying the Hamilton-Jacobi equation

$$g^{\mu\nu}(\partial_\mu S - qA_\mu)(\partial_\nu S - qA_\nu) = m^2. \qquad (4.123)$$

[This semiclassical propagator is just a function of the spacetime co-ordinates and is, of course, quite different from K.]. The Hamilton-Jacobi equation is solved formally by the function

$$S = S_0(x) + q\int_x^{x'} A_\mu \mathrm{d}x^\mu \qquad (4.124)$$

where S_0 satisfies the Hamilton-Jacobi equation in the absence of the electromganetic field. Therefore the probability amplitude for propagation from x to x', in the semiclassical limit, is now controlled by the factor

$$P = \exp\left[+iq\int_x^{x'} A_\mu \mathrm{d}x^\mu\right]. \qquad (4.125)$$

If x and x' lie on different sides of the singular surface we have no means of even defining this amplitude.

Unless some external criterion is given, defining the integral, we cannot proceed any further. In the example given above, no such natural criterion is available and we have abandoned the problem as ill-defined. But such a criterion arises in a disguised form in some other situations when the function f depends on time t. This is because the calculations involving path integrals are usually performed in Euclidean space (or done in the Lorentzian space with a suitable $i\varepsilon$ prescription, which is equivalent) and the final result is analytically continued back to Lorentzian space via the rule $t_{\text{Euclidean}} \equiv t_E = it$; $s_E = is$. [Under this substitution we induce the change

$$\exp\left[-i\int d\tau(\dot{t}^2 - \dot{\mathbf{x}}^2)\right] \to \exp\left[-\int d\tau_E(\dot{t}_E^2 \to \dot{\mathbf{x}}^2)\right] \tag{4.126}$$

making the argument of the exponent negative definite.] *The nature of the singularities of the function $f(x)$- and $A_\mu(x)$- can change drastically under Euclidean continuation.* Some potentials which are singular in Lorentz spacetime can acquire strange interpretations in the Euclidean sector and could give raise to interesting effects.

As an example, consider an A_μ which is obtained by replacing y by t in (4.119):

$$A_\mu = B\left(\frac{x}{x^2 - t^2}, -\frac{t}{x^2 - t^2}, 0, 0\right) \tag{4.127}$$

which corresponds to the gauge-function

$$f = \begin{cases} B\tanh^{-1}(t/x); & |t| < |x| \\ B\tanh^{-1}(x/t); & |t| > |x| \end{cases}. \tag{4.128}$$

Everything we said above for the potential in (4.119) is applicable to this case as well; the integral I is ill-defined for paths which across the $x = \pm t$ planes (light-'cone's). Unless an extra prescription is given, the integral cannot be evaluated in the Lorentzian space.

Suppose we now decide to invoke the Euclidean prescription to evaluate the path integrals. This would have made no difference in the case of (4.119); that potential is as singular in the Euclidean sector as in the Lorentzian sector. However, the situation is different for the potential in (4.127). The singularity in two planes ($x = \pm t$) in the Lorentzian space collapses to a singularity along a hypersurface in the Euclidean sector. [This is most easily seen in the xt plane. The function in (4.127) is singular on the two lines $x = \pm t$ in the Lorentzian sector. Euclidean continuation will change the denominator of (4.127) to $(t_E^2 + x^2)$; so the potential in the Euclidean sector will only be singular at the origin of the $t_E x$-plane. This is precisely the kind of situation we have encountered previously in the case of Aharonov-Bohm potential which we know how to handle.] In the Euclidean sector the electromagnetic coupling term becomes:

$$\begin{aligned} \exp\left[iq\int d\tau\, A_\mu \dot{x}^\mu\right] &= \exp iqB\int\left[\frac{x\mathrm{d}t - t\mathrm{d}x}{x^2 - t^2}\right] \\ &\to \exp qB\int\left[\frac{x\mathrm{d}t_E - t_E\mathrm{d}x}{x^2 + t_E^2}\right] \\ &= \exp\left[qB\int d\left(\tan^{-1}\frac{t_E}{x}\right)\right]. \end{aligned} \tag{4.129}$$

Note that the argument of the exponent is real and of indefinite sign but is perfectly bounded. We can therefore define $A_\mu^{(E)}$ through the relation

$$A_\mu^{(E)}\, \mathrm{d}x_{(E)}^\mu = qB\, \mathrm{d}\left(\tan^{-1}\left(\frac{t_E}{x}\right)\right) \qquad (4.130)$$

corresponding to

$$f_{(E)}(x_{(E)}^i) = B\theta_E = B \tan^{-1}\left(\frac{t_E}{x}\right). \qquad 4.131)$$

This is similar to the Euclidean version of Aharonov-Bohm potential in the $t_E x$-plane [In the Euclidean space, $t_E x$-plane is treated in the same footing as xy-plane]. Since we know how to handle the singularity at the origin, we can now compute the kernel etc. by this method and analytically continue back to the Lorentzian spacetime. If the kernel and Green's function are defined in the Euclidean sector, they will exhibit periodicity in the θ_E. On continuing back to Lorentzian sector, we will get a Green's function which is invariant under the strange-looking transformation

$$\tan^{-1}\left(\frac{it}{x}\right) \to \tan^{-1}\left(\frac{it}{x}\right) + \frac{2\pi}{B}. \qquad (4.132)$$

By this process, we *can*, if we want, give meaning to an object which was originally ill-defined in the Lorentzian sector. We have to, however, pay a price for this luxury, which is illustrated in figure 1. Let us compare the propagation amplitude for a particle to go from A to B with the amplitude to propagate from B to A. Neither amplitude can be calculated in the Lorentz spacetime because of the singular surface coming in the way. Calculating it using the Euclidean extension is illustrated diagrammatically in figure 1. To go from A to B, we first go from A to C in the *Euclidean* space and then proceed from C to B via analytic continuation. Now consider the amplitude for propagation from B to A. Our procedure will be to go along the path BC first (analytic continuation) followed by CA (in Euclidean space). But in the Euclidean space, the paths A to C and C to A will give amplitudes which differ by

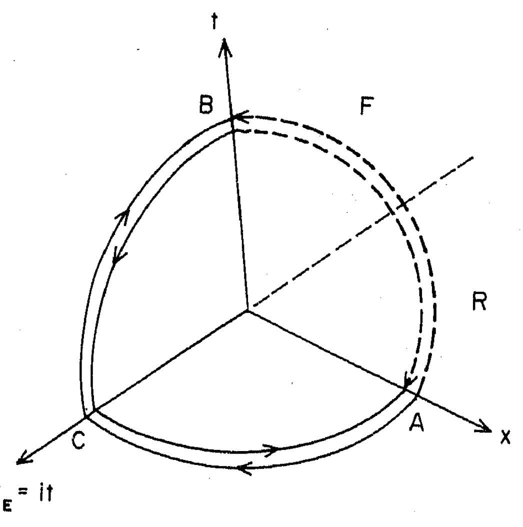

Figure 1.

a *real* factor in the exponent! The angular co-ordinate in the Euclidean xt_E-plane changes from 0 to $(\pi/2)$ as we go along A to C and contributes a term $\exp qB(\pi/2)$ (with no i-factor; see (4.130).); the path from C to A gives $\exp - qB(\pi/2)$ so that these two amplitudes differ by the factor $\exp qB\pi$ making the probabilities differ as

$$\mathscr{P}(A \to B) = \mathscr{P}(B \to A)\exp + 2\pi qB. \tag{4.133}$$

Purely from the mathematical point of view, this result suggests that it is less probable for the region $\mathscr{R}$ to gain particle from $\mathscr{F}$ than to lose particles to $\mathscr{F}$ if $qB > 0$. (For antiparticles with opposite charge, the signs have to be changed.) If we compute physical processes in $\mathscr{R}$ using the usual Feynman rules, then one has to add external sources along the boundaries $x = \pm t$ to take this effect into account (for a discussion of this point, see Lee 1986).

The essential idea which allowed one to perform the above calculation is the fact that *singularities on the light cone* are 'regularized' in the Euclidean sector. (So, obviously, the trick works only for a very special class of singular gauges.). The above result can be seen more clearly by noting that, in the Lorentzian sector, our prescription is equivalent to replacing the quantity $(x^2 - t^2)$ by $(x^2 - t^2 + i\varepsilon)$. This has the consequence of giving an imaginary part to the integral

$$J = \int_x^y A_i \,\mathrm{d}x^i = B\left[\int_{\mathscr{P}}^{\mathscr{P}'} \frac{x\mathrm{d}t - t\mathrm{d}x}{x^2 - t^2 + i\varepsilon}\right] \tag{4.134}$$

when the path crosses the singular surface. For example, if we evaluate J along a straight line of infinitesimal length: $x = s + c$; $t = -s + c$; $-\delta < s < +\delta$ (which cuts the singular line $x = t$ orthogonally at (c, c) and has the length 2δ) the integral will pick up, over and above the principal value, the imaginary part

$$\operatorname{Im} J = -\tfrac{1}{2}\pi B \tag{4.135}$$

where we have taken $\mathscr{P}$ and $\mathscr{P}'$ in the right and the top quadrants respectively. The amplitude to propagate from $\mathscr{P}'$ to $\mathscr{P}$ will contain the negative of the above result in the exponent. *The imaginary part implies that these two amplitudes differ by more than a phase:*

$$|\mathscr{A}(\mathscr{P} \to \mathscr{P}')|^2 = |\mathscr{A}(\mathscr{P}' \to \mathscr{P})|^2 \exp(+2\pi qB). \tag{4.136}$$

What do we make of this strange result? In a way it is quite upsetting especially since there is no electromagnetic field anywhere in sight. The initial impulse, of course, will be to suspect that there is something wrong in the way we have made our Euclidean continuation. Probably we should search for some other procedure which will only contribute a phase when we go from A to C or C to A. It is doubtful whether one can come up with such a procedure; but even if one does for this particular gauge, it will be possible to produce some other external gauge in which crazy behaviour occurs. (Notice that, we have just written down the Lorentzian action, and analytically continued it in the most straightforward way; if the gauge was non-singular, no one would have raised eyebrows at this procedure.).

On second thoughts, it will be clear that one is not quite right in assuming that the electromagnetic field is zero. The easiest way to see this is again to look at the

Euclidean sector. Here the vector potential corresponds to that of Aharonov-Bohm cylinder situated at the origin of the $t_E x$-plane and hence contributes a delta function field at the origin. Analytically continuing back to the Lorentz spacetime, one would be led to a delta function like field along the null surfaces. Such a field will, of course, make the particles and antiparticles to be polarized away from the singular surfaces $x = \pm t$ in the manner discussed above. So if we take Euclidean interpretation seriously, then we should be cautious about deciding which potentials represent 'pure' gauge.

Notice, however, that there exists one *physical* reason for treating the above result as of without any operational significance, in the context of electromagnetism. This is because we do not know how to give operational meaning to the condition: 'Describe a electromagnetic field in a particular gauge corresponding to a given function $A_i(x)$'. All we can do with capacitor plates and coils of wires is to produce a given electromagnetic *field*; there is no known way of *operationally implementing* a particular *gauge*. So if we obtain a strange result in a singular gauge, all we can do is to invoke a rule that such gauges should not be used in the mathematical description. [Presumably, this issue is related to the canonical implementability of the gauge transformation mentioned at the end of §4.6].

All the same there is an important lesson to be learnt from the above example, which motivated its study in the first place. It turns out that the corresponding situation in the case of gravity has an entirely different physical meaning. Since it is possible to invoke observers in different states of motion, it is probable that one can give a physical realization to the choice of different co-ordinate systems. Such a choice is analogous to the choice of a gauge in electrodynamics. We will see that the analysis performed above in electrodynamics, when carried out in some spacetime manifolds *described in singular co-ordinate charts* leads to the familiar 'thermal' effects in gravity.

5. Quantum theory in external gravitational field

We shall now consider the analogies between quantum theory in an external electromagnetic field and quantum theory in a gravitational field. The discussion will be limited to three important features: (1) The formal correspondence between pair creation in an electric field and pair creation in cosmology. (2) A comparison between gauge invariance in electromagnetism and coordinate invariance in general relativity. (3) The thermal effects which arise due to quantization in singular gauges.

5.1 *Pair creation in electric field and expanding universe*

There is a formal correspondence between pair creation in a time dependent electric field and pair creation in an expanding Friedmann universe. This can be seen as follows: Consider, for example, the action for a scalar field Φ

$$A = -\int d^4x\sqrt{-g}\,\tfrac{1}{2}\Phi[\Box + m^2 + \tfrac{1}{6}R]\Phi \tag{5.1}$$

in the Friedmann spacetime with the line element

$$ds^2 = a^2(t)(dt^2 - dr^2). \tag{5.2}$$

[This action is conformally invariant in the limit of m going to zero. This fact makes our analysis easy, however the results are valid even for non-conformal coupling]. Writing Φ as (ϕ/a) and exploiting the conformal flatness of the metric, we can reduce the action to the form

$$A = -\frac{1}{2}\int d^4 x \phi[\Box_{\rm flat} + m^2 a^2(t)]\phi. \tag{5.3}$$

To study the pair creation, we can again use the effective lagrangian method. The kernel we need is

$$\begin{aligned} K(x, y; s) &= \langle x|\exp[-is\tfrac{1}{2}[\Box + m^2 a^2(t) - i\varepsilon]]|y\rangle \\ &= \int \frac{d^3 p}{(2\pi)^3}\langle t|\exp[-is\tfrac{1}{2}[\partial_t^2 + p^2 + m^2 a^2(t) - i\varepsilon]]|t'\rangle \\ &= \int \frac{d^3 p}{(2\pi)^3}\exp[-is\tfrac{1}{2}(p^2 - i\varepsilon)]\mathscr{G}(t, t'; s) \end{aligned} \tag{5.4}$$

where $\mathscr{G}$ is the propagator for the quantum mechanical hamiltonian

$$H = -\frac{1}{2}\frac{\partial^2}{\partial t^2} - \tfrac{1}{2}m^2 a^2(t). \tag{5.5}$$

Comparing this expression with the corresponding one for the electric field we can make the identification: $m^2 a^2(t) \Leftrightarrow (p_z - qA(t))^2$. Thus there exists an one-to-one correspondence between time dependent electric fields and expanding Friedmann universes as far as the quantization of an external scalar field is concerned.

As an example, consider the case of constant electric field. The analogue in cosmology will be a universe with the conformal factor:

$$a^2(t) = \frac{1}{m^2}(p_z + qEt)^2 \equiv \alpha^2(t + t_0)^2; \; \alpha = \left(\frac{qE}{m}\right). \tag{5.6}$$

In the more familiar coordinate system with

$$ds^2 = d\tau^2 - a^2(\tau)\mathbf{dr}^2 \tag{5.7}$$

this corresponds to the expansion law

$$a(\tau) = (2\alpha\tau)^{1/2} \propto \tau^{1/2} \tag{5.8}$$

This corresponds to a radiation-dominated universe. Similar correspondences can be established in other cases which allow one to translate the results in one physical situation to another.

5.2 *Quantum theory in a Milne universe*

We shall now take up the analogy between gauge and coordinate invariance (Padmanabhan 1990). To do this, one requires some region of spacetime manifold which can be represented conveniently in two different co-ordinate systems; one system in which the metric is static and another in which it depends only on time. The simplest choice happens to be the upper quarter of the flat spacetime. In this 'top-quarter' of the Minkowski spacetime (i.e. the region $T>|X|$ which we will call (U'), the line element can be expressed in two different ways:

$$\begin{aligned} ds^2 &= dT^2 - dX^2 - dY^2 - dZ^2 \\ &= d\rho^2 - g^2\rho^2 dl^2 - dY^2 - dZ^2 \\ &= [\exp(2gt)](dt^2 - dx^2) - dY^2 - dZ^2 \end{aligned} \tag{5.9}$$

by the transformation $gX = \exp(gt)\sinh gx$ and $gT = \exp(gt)\cosh gx$; $g\rho = \exp(gt)$. The intermediate form of the transformation shows that the metric belongs to the class of anisotropically expanding cosmological solutions. Because of this similarity, we will call this co-ordinate system the 'Milne Universe.' Since the 'static gauge' now is just the inertial co-ordinates, we only have to work out the quantum theory in the Milne universe. We want to study the evolution of a quantum field along the hypersurfaces defined by constant-t (in the Milne coordinates) and compare it with the conventional Minkowski quantization. Since the metric in Milne co-ordinates depend only on t, the Klein-Gordon equation

$$\frac{1}{\sqrt{-g}}\partial_i(\sqrt{-g}\,g^{ik}\partial_k\phi) - m^2\phi = 0 \tag{5.10}$$

can be separated as

$$\phi(\mathbf{x},t) = \sum_{\mathbf{x}} \{a_k f_k(t)\exp(i\mathbf{k}.\mathbf{x}) + \text{h·c}\} = \sum_{\mathbf{k}} (\phi_k^{(+)} + \text{h·c}). \tag{5.11}$$

This equation has the two linearly independent solutions which may be taken to be $H_{ip}^{(1)}(\rho)$ and $H_{ip}^{(2)}(\rho)$ where $H_\mu(z)$ is the Hankel function. We write

$$\begin{aligned} f(t) &= c_1 H_{ip}^{(1)}(\rho) + c_2 H_{ip}^{(2)}(\rho) \\ &= c_1(J_\nu + iN_\nu) + c_2(J_\nu - iN_c) \\ &= (c_1 + c_2)J_\nu(\rho) + i(c_1 - c_2)N_\nu(\rho) \\ &\equiv b_1 J_\nu(\rho) + ib_2 N_\nu \end{aligned} \tag{5.12}$$

where $\nu = ip$ and J_ν and N_ν are the Bessel functions (Gradshteyn and Ryzhik 1965). We are interested in the limits $t \to \pm\infty$. From the properties of the Bessel functions, it is easy to see that

$$\lim_{t\to-\infty} f_k(t) \cong \left\{(b_1 + ib_2\cot\nu\pi)\frac{\rho^\nu}{2^\nu}\frac{1}{\Gamma(1+\nu)} - ib_2\operatorname{cosec}\nu\pi\cdot\frac{\rho^{-\nu}}{2^{-\nu}}\frac{1}{\Gamma(1-\nu)}\right\}. \tag{5.13}$$

Since $\rho^{\pm\nu} \simeq \exp(\pm g\nu t) = \exp(\pm i|k_x|t)$ (in U) and we want f to go as $\exp(-i\omega t)$ for the positive frequency mode only the second term is admissible. Therefore the positive frequency modes in the infinite past are the ones obtained by the condition

$$b_1 = -ib_2\frac{\cos ip\pi}{\sin ip\pi} = -ib_2\frac{\cosh p\pi}{i\sin \rho\pi} = -b_2\coth\pi p. \tag{5.14}$$

We should take $f_k(t)$ to be

$$f_k(t) = ib_2 \operatorname{cosec} \nu\pi J_{-\nu} = -\frac{b_2}{\sinh p\pi} J_{-\nu}. \tag{5.15}$$

The value of b_2 is fixed by the normalization condition:

$$i(f^*\dot{f} - f\dot{f}^*) = (2\pi)^{-3}. \tag{5.16}$$

Straightforward calculation gives,

$$|b_2|^2 = \left(\frac{1}{2\pi}\right)^3 \frac{\pi(\sinh p\pi)}{2g} = \left(\frac{1}{2\pi}\right)^3 \left(\frac{\pi}{2g}\right)(\sinh p\pi) \tag{5.17}$$

so that

$$|b_1|^2 = |b_2|^2 \coth^2 \pi p = \left(\frac{1}{2\pi}\right)^3 \left(\frac{\pi}{2g}\right) \frac{\cosh^2 p\pi}{\sinh p\pi}. \tag{5.18}$$

We know now that, the solution which behaves as $\exp(-i\omega t)$ near $t \to -\infty$ in U, is

$$\begin{aligned} f(t) &= -\frac{b_2}{\sinh p\pi} J_{-\nu}(\rho) \\ &= -\left(\frac{1}{2\pi}\right)^{3/2} \left(\frac{\pi}{2g}\right)^{1/2} \frac{J_{-\nu}(\rho)}{\sinh^{1/2} p\pi}. \end{aligned} \tag{5.19}$$

It is clear from the asymptotic form of the equation for f that we will not get $\exp(\pm i\omega t)$ in the infinite future. Therefore the positive frequency mode has to be identified by the WKB analysis, as in the electromagnetic field. This analysis shows that the proper mode is the one which behaves as $\exp(-i\rho)$ in the infinite future. Since $H_\nu^{(2)}$ behaves as $\exp(-i\rho)$ near large ρ we can set $c_1 = 0$ and take the solution to be

$$g(t) = c_2 H_\nu^{(2)}(\rho). \tag{5.20}$$

To normalize this solution, we will again use the condition

$$W = i(g^*\dot{g} - g\dot{g}^*) = (2\pi)^{-3}. \tag{5.21}$$

This gives

$$|c_2|^2 = \left(\frac{1}{2\pi}\right)^3 \left(\frac{\pi}{4g}\right) \exp(+p\pi). \tag{5.22}$$

We can now express the positive frequency solution of the infinite past in terms of the positive and negative frequency solutions of the infinite future and identify the Bogoliubov coefficients. Using the identities

$$[\exp(i\nu\pi)] H_\nu^{(1)} = \frac{i}{\sin \pi\nu} J_\nu - \frac{i}{\sin \pi\nu} [\exp(i\nu\pi)] J_{-\nu}$$

$$[\exp(-i\nu\pi)] H_\nu^{(2)} = -\frac{i}{\sin \nu\pi} J_\nu + \frac{i}{\sin \pi\nu} [\exp(-i\nu\pi)] J_{-\nu} \tag{5.23}$$

and

$$[\exp(i\nu\pi)] H_\nu^{(2)} + [\exp(-i\nu\pi)] H_\nu^{(2)} = 2J_{-\nu} \tag{5.24}$$

it is easy to show that

$$f(t)=\frac{1}{\sqrt{2}}\left(\frac{1}{\sinh \pi p}\right)^{1/2}\{[\exp(-\pi p/2)]g^*(t)+[\exp(\pi p/2)]g(t)\}. \tag{5.25}$$

This corresponds to the Bogoliubov coefficients

$$\alpha=\frac{1}{\sqrt{2}}\left(\frac{1}{\sinh \pi p}\right)^{1/2}\exp(\pi p/2);\ \beta=\frac{1}{\sqrt{2}}\left(\frac{1}{\sinh \pi p}\right)^{1/2}\exp(-\pi p/2). \tag{5.26}$$

We see that

$$\alpha^2-\beta^2=\frac{1}{2}\frac{[\exp(\pi p)-\exp(-\pi p)]}{\sinh \pi p}=1 \tag{5.27}$$

as it should. The number density of created particles is

$$\beta^2=-\frac{1}{2}\frac{2\exp(-\pi p)}{\exp(\pi p)-\exp(-\pi p)}=\frac{1}{[\exp(2\pi p)]-1}=\frac{1}{[\exp((2\pi/g)|k_x|)]-1} \tag{5.28}$$

which corresponds to a thermal spectrum of particles in the longitudinal momentum with the temperature $(g/2\pi)$.

Similar result can be obtained in a different context which is well-known in literature (Chitre and Hartle 1977). Investigations in the study of particle production by expanding Friedmann universes have shown that, in a spatially flat model with the expansion law $a(t)\propto t$ a thermal spectrum of particles is produced at late times. Interestingly enough, the analysis is valid in 2-dimension as well in which case the spacetime is just the TX-sector of the Milne universe. Our analysis shows that the other two dimensions merely go for a ride.

Even though (5.28) corresponds to a temperature of $(g/2\pi)$, the result is very different from the standard result obtained in Rindler frame (Fulling 1973; Davies 1975; Unruh 1976) for two reasons: (i) We are working in the upper and lower quarters, while the Rindler co-ordinates exist only in the right and left quarters. This makes the entire situation quite different. (ii) There is no 'particle creation' in the Rindler co-ordinates (τ', ρ, y, z). The Rindler mode functions behave as $\exp \pm i\omega\tau'$ for all times. The conventional result only says that these mode functions are connected to the Minkowski modes by a Bogoliubov transformation with off-diagonal term which leads to a result similar to that in (5.28). In contrast, we are now working with a non-static background; the positive frequency mode in the infinite past *does* get mixed up with positive *and* negative frquency modes of the infinite future.

The last feature is somewhat disturbing; it shows that, if we had no prior knowledge that we are dealing with flat spacetime we would have accepted the result in (5.28) as 'genuine' particle creation! In fact, the procedure we have followed is identical to the one usually adopted to study field theory in expanding universes. Our result suggests that 'particle creation' can be spurious effect even in a curved spacetime; it certainly will be co-ordinate dependent. We have to produce a sensible criterion which will distinguish particle creation due to spacetime curvature from effects due to the choice of co-ordinates. Without such a criterion, it is meaningless to talk of quantum field theory in curved space.

The same region of spacetime can also be represented in the static Minkowski coordinates with the modes $\Phi^{\pm} \sim \exp \mp i(\omega T - \mathbf{K}\cdot\mathbf{X})$. In this gauge, of course, there is no particle creation and the vacuum state remains a vacuum state at all times. It follows that the particle concepts in the Minkowski and Milne co-ordinates are inequivalent. The Minkowski vacuum will contain Milne particles. We can obtain this particle content by evaluating the scalar products (Φ^-, f^+) and (Φ^-, g^+). It turns out that

$$|(\Phi^-, f^+)|^2 = \frac{1}{\exp[(2\pi|k_x|)/g] - 1}; \ |(\Phi^-, g^+)|^2 = 0 \tag{5.29}$$

In other words, the particle definitions in the Minkowski and Milne co-ordinates agree in the asymptotic future. However, the Minkowski vacuum will contain a thermal spectrum of particles in the asymptotic past.

Similar co-ordinate system can be introduced in the 'bottom-quarter' $(-T > |X|)$ of the Minkowski spacetime. In this region, the particle concepts match in the asymptotic past but not in the future.

5.3 *Spacetime manifold in singular gauges*

We shall next consider the gravitational analogue of the effects discussed in § 4.7. It turns out that the 'thermal' effects in certain spacetime provide this analogy (Padmanabhan 1991a). Consider a patch of spacetime, which, in suitable coordinate system, has the line element,

$$\begin{aligned} ds^2 &= + B(r)dt^2 - B^{-1}(r)dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2) \\ &\equiv + B(r)dt^2 - B^{-1}(r)dr^2 - dL^2 \end{aligned} \tag{5.30}$$

or

$$\begin{aligned} ds^2 &= B(x)dt^2 - B^{-1}(x)dx^2 - dy^2 - dz^2 \\ &\equiv B(x)dt^2 - B^{-1}(x)dx^2 - dL^2. \end{aligned} \tag{5.31}$$

Co-ordinate systems of the form (5.30) can be introduced in parts of Schwarschild and de Sitter spacetimes while the choice $[B(x) = 1 + 2gx]$ in (5.31) represents a uniformly accelerated frame (Rindler frame) in flat spacetime. We will be concerned basically with the structure of the metric in the rt or xt plane. Since this structure is essentially the same in both (5.30) and (5.31) we shall work throughout with (5.30); the results are extendable to (5.31) in a straightforward manner.

The exponent of the kernel $K(x', x; s)$ will now contain the integral

$$\mathscr{A} = \int_0^s d\tau g_{ik}\dot{x}^i\dot{x}^k = \int_0^s d\tau[B\dot{t}^2 - B^{-1}\dot{x}^2 + \ldots]. \tag{5.32}$$

Quite obviously we will run into problems if B vanishes along the path of integration. From the nature of our metric it is easy to see that the surfaces on which B vanishes are null surfaces corresponding to infinite redshift ('horizons').

To study the effect of horizons, let us proceed in the following manner: Suppose that at some $r = r_0(> 0)$, $B(r)$ vanishes, $B'(r)$ finite and nonzero. Then near $r = r_0$, we can expand $B(r)$ as,

$$B(r) = B'(r_0)(r - r_0) + \mathcal{O}[(r - r_0)^2]$$
$$\equiv R(r_0)(r - r_0). \tag{5.33}$$

As long as the points 1 and 2 (between which the transition amplitude is calculated) are in the same side of the horizon [i.e. both are at $r > r_0$ or both at $r < r_0$] the integral in the action is well defined and real. But if the points are located at two sides of the horizon then the integral does not exist due to the divergence of $B^{-1}(r)$ at $r = r_0$.

Let us first review briefly the conventional derivation of thermal effects using path integrals, say, in the context of Schwarschild blackhole. Given the co-ordinate system of (5.30), in some region R, we first verify that there is no *physical* singularity at the horizon. Having done that, we extend the geodesics into the past and future and arrive at two further regions of the manifold not originally covered by the co-ordinate system in (5.30). Let us label these regions as F and P. It is now possible to show that the probability for a particle with energy E to be lost from the region R (i.e. probability for propagation from $\mathscr{P}$ to $\mathscr{P}'$) in relation to the probability for a particle with energy E to be gained by the region R (i.e. propagation from $\mathscr{P}'$ to $\mathscr{P}$) by the equation

$$P_{(\text{loss})} = P_{(\text{gain})} \exp - \beta E. \tag{5.34}$$

This is equivalent to assuming that the region R is bathed in radiation at temperature β^{-1}. [The derivation in e.g. Gibbons and Hawking (1976) actually relates the amplitudes involving past and future horizons; but it can be reexpressed in the above manner.]

The above result can be interpreted differently, so as to bring out the connection with the case of electromagnetic field. This is most easily done by considering the semiclassical approximation to the path integral propagator, expressed in the saddle point approximation, as:

$$\mathscr{G}(x_2, t_2; x_1, t_1) = \mathscr{G}(2, 1) = N \exp iA(2, 1) \tag{5.35}$$

where A is the action functional satisfying the classical Hamilton-Jacobi equation. For a particle of mass m, moving in our spacetime the Hamilton-Jacobi equation will be:

$$g^{ik}(\partial_i A)(\partial_k A) - m^2 = 0. \tag{5.36}$$

The solution to this equation can be represented as

$$A = -Et + J\theta + A_r(r) \tag{5.37}$$

with,

$$A_r(r) = \pm \int^r dr B^{-1}(r)[E^2 - B(r)(m^2 + J^2/r^2)]^{1/2}. \tag{5.38}$$

The sign ambiguity of the square-root is related to the "outgoing" $[(\partial A/\partial r) > 0]$ or "ingoing" $[(\partial A/\partial r) < 0]$ nature of the particle.

The result above is similar to the one obtained earlier in (4.122). This semiclassical amplitude can be computed using this expression only if the integral in this expression exist; and it does not exist if the two points lie on different sides of the horizon.

Thus, in order to obtain the probability amplitude in (5.35) for crossing the horizon (i.e. when 1 and 2 are on two sides of the horizon), we have to give some extra prescription for evaluating the integral. Since the surface $B=0$ is null (just like $x=\pm t$ in the electrodynamic case) we may carry out the calculation in the Euclidean space or—equivalently—use the $i\varepsilon$ prescription to specify the contour over which the integral has to be performed around $r=r_0$. The usual $i\varepsilon$ prescription can be easily shown to imply that we should take the contour for defining the integral to be an infinitesimal semi-circle above the pole at $r=r_0$. Thus, the contour is along the real line from, say, $r_1(0<r_1<r_0)$ to $(r_0-\varepsilon)$ and from $(r_0+\varepsilon)$ to, say, $r_2(r_2>r_0)$. From $(r_0-\varepsilon)$ to $(r_0+\varepsilon)$ we go along a semicircle of radius ε in the upper complex plane.

Consider an outgoing particle $[(\partial A/\partial r)>0]$ at $r=r_1<r_0$. What is the amplitude for it to cross the horizon? Clearly, the contribution to A in the range $(r_1, r_0-\varepsilon)$ and $(r_0+\varepsilon, r_2)$ is real. Therefore,

$$A(\text{outgoing}) = -\int_{r_0-\varepsilon}^{r_0+\varepsilon} (\mathrm{d}r/B(r))[E^2 - B(m^2+J^2/r^2)]^{1/2} + (\text{real part}). \tag{5.39}$$

[The minus sign corresponds to the initial condition that $(\partial A/\partial r)>0$ at $r=r_1<r_0$. For the sake of definiteness we have assumed R in (5.33) to be positive, so that $B<0$, at $r<r_0$. For the cases with $R<0$, the answer has to be modified by a sign change.] Evaluating the integral, in the limit of $(\varepsilon\to 0)$, we get

$$\begin{aligned} A(\text{outgoing}) &= -[E/R(r_0)](-i\pi) + (\text{real part}) \\ &= [i\pi E/R(r_0)] + (\text{real part}). \end{aligned} \tag{5.40}$$

Now consider an ingoing particle $[(\partial A/\partial r)<0]$ at $r=r_2>r_0$. The corresponding action is,

$$\begin{aligned} A(\text{ingoing}) &= -\int_{r_0+\varepsilon}^{r_0-\varepsilon} (\mathrm{d}r/B(r))\cdot[E^2 - B(m^2+J^2/r^2)]^2 + (\text{real part}) \\ &= [E/R(r_0)](+i\pi) + (\text{real part}) \\ &= [-i\pi E/R(r_0)] + (\text{real part}). \end{aligned} \tag{5.41}$$

Taking the modulus to obtain the probability, we get,

$$P(\text{outgoing}) = N\exp[-2\pi E/R(r_0)] \tag{5.42}$$

and

$$P(\text{ingoing}) = N\exp[+2\pi E/R(r_0)] \tag{5.43}$$

so that

$$P(out) = \exp[-4\pi E/R(r_0)]\cdot P(in). \tag{5.44}$$

This result is quite similar to our earlier result (4.136) in the case of electromagnetic field, and shows that it is more likely for a particular region to gain particles than lose them. If one tries to do a consistent quantum field theory in this region, one has to introduce source terms at the singular boundaries. (a detailed discussion of this point can be found in Lee (1986). Further, the exponential dependence on the energy allows one to give a 'thermal' interpretation to this result. In a system with temperature β^{-1} then the absorption and emission probabilities are related by

$$P[\text{emission}] = \exp(-\beta E)P[\text{absorption}]. \tag{5.45}$$

Comparing (5.44) and (5.45) we identify the temperature of the horizon in terms of $R(r_0)$. Equation (5.39) is based on the assumption that $R > 0$. [see the comment after (5.39)]. For $R < 0$ there will be a change of sign in this equation. Incorporating both the cases, we can write the general formula for horizon temperature to be

$$\beta^{-1} = |R|/4\pi \tag{5.46}$$

For the Schwarschild blackhole,

$$B(r) = (1 - 2M/r) \approx (1/2M)(r - 2M) + \mathcal{O}[(r - 2M)^2] \tag{5.47}$$

giving $R = (2M)^{-1}$, and the temperature:

$$\beta^{-1} = |R|/4\pi = 1/8\pi M. \tag{5.48}$$

[The following point is worth noting regarding the derivation of the thermal effect in the case of a Schwarschild blackhole: The regularization procedure which is adopted above is equivalent to replacing M by $(M - i\varepsilon)$, where M is the mass of the blackhole. This is identical to the regularization procedure which would have been adopted in standard field theory if one is dealing with particles of mass M. Probably this result has no much significance, but it certainly appears as an interesting coincidence.]. For the de Sitter spacetime,

$$B(r) = (1 - H^2r^2) = 2H(H^{-1} - r) = -2H(r - H^{-1}) \tag{5.49}$$

giving

$$\beta^{-1} = |R|/4\pi = H/2\pi. \tag{5.50}$$

Similarly for a metric of the uniformly accelerated frame

$$B(x) = (1 + 2gx) = 2g[x + (2g)^{-1}] \tag{5.51}$$

and

$$\beta^{-1} = (g/2\pi). \tag{5.52}$$

The formula can be used for more complicated metrics as well, and gives the same results as obtained by more detailed methods.

The above analysis is *not* intended to be a derivation of the thermal effects; rather, it is *interpretation* of results derived by more rigorous methods. This interpretation, however, has the advantage that it allows one to obtain the thermal effects by invoking a simple prescription for handling the integrals across the horizon and emphasizes the role played by the singular gauge.

The Rindler frame discussed above is usually considered to be part of field-free region, i.e. it represents flat spacetime in a curvilinear co-ordinate system. Our earlier discussion on pure gauge potential suggests that this aspect needs to be looked at closely, especially if Euclidean continuations are used to interpret the theory. In fact, it has been pointed out by Christensen and Duff (1978) that the Euler characteristic of the Euclidean sector—obtained by analytically continuing in the Rindler time co-ordinate—is different from that of standard (Euclidean) space. The difference arises precisely due to the nature of the singularities along the light cone in the Rindler gauge.

6. Conclusions

One of our aims is to compare the gauge invariance of the particle concept (in electromagnetism) with the co-ordinate invariance of the particle concept (in gravity). To do this, we studied the particle definition in two different gauges (which represent a constant electric field) in Part 4 and the particle definition in two different co-ordinate systems (which represent a part of flat spacetime) in Part 5. Let us now compare the results.

To begin with, we need to clarify some conceptual issues. From a mathematical point of view, the curvature tensor plays the role in gravity which is analogous to that of the field tensor F_{ik} in electromagnetism. (The vanishing of curvature, for example, signals the absence of gravitational field just as the vanishing of the field tensor signals the absence of the electromagnetic field.) Similarly we may set up a mathematical correspondence between the vector potential and the affine connection. These constructs, of course, are not unique; a given curvature tensor can be obtained from different sets of connections just as a given electromagnetic field can be obtained from different sets of vector potential. A gauge transformation connects the different choices in electromagnetism and the choice of co-ordinate systems connects the different choices in the case of gravity.

While such a mathematical parallelism can be set up between the two there are certain important operational differences. A given electromagnetic field will be produced in the laboratory by using, say, capacitor plates and coils of wire. Such a field will necessarily be bounded in space and time. More importantly, we do not know any operational means of implementing a particular gauge potential to describe the field. (Capacitor plates and coils of wire do not dictate a gauge choice.) Thus the vector potential is strictly unobservable in electromagnetism (Even phenomenon like Aharonov-Bohm effect only measure the flux and not the value of the gauge potential). By asking an operationally well-defined question in electromagnetism, we can be sure that the results will be independent of the gauge.

The situation is far more unclear in gravity. It is probably possible to produce a particular curvature in spacetime at one instant by a suitable arrangement of masses, say. At subsequent moments, the system will evolve in a particular way based on the laws of gravity. [Coils of wire can be held in place by non-electromagnetic forces; we cannot do the similar thing for the gravitational case because any agency invoked to do it will contribute to gravity and change the field significantly.]. Let us assume that we accept this situation and try to study the quantum theory in such a background spacetime. We then face the second difficulty: Unlike in electromagnetism, we have no guarantee that the curvature will be confined to a finite region in space and time. One of the most interesting situations we want to study involves the expanding universe which certainly does not seem to have asymptotically flat spacetime regions.

There is also some differences between electromagnetism and gravity as regards the observability of the gauge choices. It is usually believed that, by using suitable rods and clocks, one can actually measure the metric of the spacetime. For example, the internal dynamics of a system in accelerated motion is expected to be governed by the proper time measured by clocks co-moving with the system. Thus observers in different states of motion may provide an operational realization of *some* of the gauge choices made in gravity. This is quite different from the situation in electromagnetism where we have no means of imposing the gauge choice.

However, physically realizable motions of observations cannot lead to gauge choice which change the behaviour of the metric at infinite distances and time; nor can it make the gauge function singular at any region. The choice of coordinate system which led to the Milne universe (or, for that matter, the more familiar Rindler coordinate system) is certainly not something which can be realized by observers moving in accordance with the laws of physics. They are mathematical constructs *just as much as a constant electric field existing from everlasting to everlasting is a mathematical construct.* This is precisely the reason we have decided to compare them in this paper.

Based on such a comparison, carried out in Parts 4 and 5, we can draw the following conclusions. If we consider quantum theory in a *strictly constant* electric field (by which we mean a field which is *never* switched off) then we do obtain results which are gauge dependent; particles are created in one gauge but not in another. Note that we cannot use the path integral technique here because no in-out vacuum states exist. It is indeed possible to reinterpret the particle creation of one gauge as an over-the-barrier-reflection in the other; such an interpretation *is* often resorted to but it does involve certain additional assumptions.

If the field is physically reasonable (in the sense that it *does* vanish in the asymptotic past and future) then one can give a gauge invariant meaning to particle creation by using the particle concepts defined asymptotically. This can be achieved quite neatly using the Schwinger's proper time method or more elaborately by matching the solutions with and without the field.

The connection with gravitational context is as follows: All the examples known in the literature in which an ambiguity in the definition of particle arises, involve co-ordinate transformations which either change the asymptotic behaviour of the metric or become singular at the 'edges' of the spacetime domain under consideration (or both). Milne and Rindler co-ordinates are similar examples of this kind. Such mathematical constructs are analogous to electric field which is never switched off. We have shown that under either circumstance the particle concept becomes ambiguous (gauge ambiguity in electromagnetism and co-ordinate dependence in gravity). Hopefully this will make the gravitational results somewhat less mysterious.

What about co-ordinate transformations which are physical in the sense that they are operationally realized by a set of observers moving in accordance with the laws of physics? A partial answer, in analogy with electromagnetism can be immediately given. If the asymptotic nature of the metric is not changed by the co-ordinate transformation then one can always provide an invariant definition of S-matrix elements using the asymptotic states; then there will be no ambiguity. A more interesting question would be to ask about the response of physical systems coupled to the quantum field when the system is in different states of motion. This has been analyzed extensively in the case of gravity. A similar study can be done in the case of electromagnetism.

Lastly one can mention a possible lesson to be learnt from this comparison. We mentioned before that the breakdown of gauge invariance in the pair creation in electromagnetic field is due to our using a gauge transformation which cannot be implemented unitarily in the Hilbert space. We suspect the results in the case of gravity are of similar origin. The co-ordinate transformations which connect Minkowski spacetime with Milne co-ordinates (or, for that matter, the Rindler co-ordinates) spoil the asymptotic behaviour of original metric. It is doubtful whether such transformations can be unitarily implemented in the Hilbert space of full quantum theory of gravity. It is very likely that quantum gravity is co-ordinate invariant (just as QED is gauge invariant) as long as one restricts oneself to transformations which are unitarily implementable. In such a full theory, the particle concept will be as much co-ordinate invariant as it is gauge invariant in QED.

The above discussion stresses the fact that the quantum theory remains meaningfully invariant only under a subset of classically allowed transformations. This subset is characterized by sensible boundary conditions at large distances. The same conclusion can be arrived at by a strictly operational approach to the problem: Any physically realizable electric field has to be confined in space and time; it can be shown that there is no ambiguity in the particle creation for such fields. Similarly, any physically realizable co-ordinate system can differ from the Minkowski co-ordinates only in a finite region of spacetime. (This excludes Milne, Rindler and a host of other co-ordinate systems as physically unrealizable.). Under such transformations, which leaves the asymptotic domain unchanged, the standard concepts of field theory will be co-ordinate invariant. We feel that the results obtained in other cases are not of practical significance.

Appendix

A more formal way of deriving the Bogoliubov coefficients in the time dependent gauge is the following: From the known asymptotic forms of D_ν functions we get, for $\tau \to +\infty$

$$D_\nu((1-i)\tau) = D_\nu\left(\sqrt{2}\exp\left(-\frac{i\pi}{4}\right)\tau\right) \approx (\sqrt{2})^\nu \tau^\nu \exp\left(-\frac{i\pi\nu}{4}\right)\exp(+\tfrac{1}{2}i\tau^2)$$

$$D_\nu^*((1+i)\tau) \cong (\sqrt{2})^{\nu^*}\tau^{\nu^*}\exp(-\tfrac{1}{2}i\tau^2) \tag{A.1}$$

while for $\tau \to -\infty$ the same functions behave as

$$\begin{aligned} D_\nu((1-i)(\exp i\pi)|\tau|) &= D_\nu\left(\sqrt{2}|\tau|\left(\exp i\frac{3\pi}{4}\right)\right) \\ &\cong (\sqrt{2})^\nu|\tau|^\nu\left(\exp i\frac{3\pi\nu}{4}\right)\exp(+\tfrac{1}{2}i\tau^2) \\ &\quad - \frac{\sqrt{2\pi}}{\Gamma(-\nu)}\exp(i\pi\nu)\exp(-\tfrac{1}{2}i\tau^2)(\sqrt{2})^{\nu^*}|\tau|^{\nu^*}\exp\left(i\frac{3\pi}{4}\nu\right) \end{aligned} \tag{A.2}$$

and

$$D_{\nu^*}((1+i)\exp(i\pi)|\tau|) = D_{\nu^*}(\sqrt{2}|\tau|\exp(-i3\pi/4))$$
$$= (\sqrt{2})^{\nu^*}|\tau|^{\nu^*}\exp\left(-\frac{3\pi\nu^*}{4}\right)\exp(-\tfrac{1}{2}i\tau^2)$$
$$-\frac{\sqrt{2\pi}}{\Gamma(-\nu^*)}\exp(-\pi\nu^*)\exp(+\tfrac{1}{2}i\tau^2)(\sqrt{2})^{\nu}|\tau|^{\nu}\exp\left(-\frac{i3\pi}{4}\nu\right). \tag{A.3}$$

Consider now the other two functions. For $\tau \to -\infty$ we have:

$$D_{\nu}(-(1-i)\tau) = D_{\nu}\left(\sqrt{2}\exp\left(\frac{i3\pi}{4}\tau\right)\right) = D_{\nu}\left(\sqrt{2}\exp\left(\frac{i7\pi}{4}|\tau|\right)\right)$$
$$= D_{\nu}\left(\sqrt{2}\exp\left(\frac{-i\pi}{4}|\tau|\right)\right.$$
$$= (\sqrt{2})^{\nu}|\tau|^{\nu}\exp\left(\frac{-i\pi\nu}{4}\right)\exp\left(\frac{i}{2}\tau^2\right)$$
$$D_{\nu^*}(-(1+i)\tau) = (\sqrt{2})^{\nu^*}\exp\left(\frac{i\pi}{4}\nu^*\right)\exp\left(\frac{-i}{2}\tau^2\right) \tag{A.4}$$

while for $\tau \to +\infty$ we get the behaviour:

$$D_{\nu}(1-i)\tau) = D_{\nu}\left(\sqrt{2}\exp\left(\frac{i3\pi\nu}{4}\right)\tau\right)$$
$$= (\sqrt{2})^{\nu}\tau^{\nu}\exp\left(\frac{i3\pi\nu}{4}\right)\exp\left(\frac{i}{2}\tau^2\right)$$
$$-\frac{\sqrt{2\pi}}{\Gamma(-\nu)}\exp(i\pi\nu)\exp(-\tfrac{1}{2}i\tau^2)\left(\sqrt{2}^{\nu^*}\tau^{\nu^*}\exp\left(\frac{i3\pi}{4}\nu^*\right)\right)$$
$$D_{\nu^*}(-(1+i)\tau) = D_{\nu^*}\left(\sqrt{2}\exp\left(\frac{-3\pi}{4}\right)\tau\right)$$
$$= (\sqrt{2})^{\nu^*}\tau^{\nu^*}\exp\left(-\frac{i3\pi\nu^*}{4}\right)\exp\left(\frac{-i}{2}\tau^2\right)$$
$$-\frac{\sqrt{2\pi}}{\Gamma(-\nu^*)}\exp(i\pi\nu^*)\exp\left(\frac{i}{2}\tau^2\right)(\sqrt{2})^{\nu}\tau^{\nu}\exp\left(-\frac{i3\pi}{4}\nu\right). \tag{A.5}$$

It is clear that near $\tau \approx -\infty$, $D_{\nu}(-(1-i)\tau)$ is the positive frequency mode while near $\tau \approx +\infty$, $D_{\nu^*}((1+i)\tau)$ is the positive frequency mode. Evolving $D_{\nu}(-(1-i)\tau)$ to

$\tau = (+\infty)$ we see

$$D_\nu(-(1-i)\tau) = -\frac{\sqrt{2\pi}}{\Gamma(-\nu)}\exp(i\pi\nu)\exp\left(\frac{i3\pi}{4}\nu^*\right)\exp\left(-\frac{\pi\nu^*}{4}\right)D_{\nu^*}((1+i)\tau)$$
$$+\exp\left(\frac{i3\pi\nu}{4}\right)\exp\left(\frac{i\pi\nu}{4}\right)D_\nu((1-i)\tau)$$
$$= -\frac{\sqrt{2\pi}}{\Gamma(-\nu)}\exp\left(\frac{i\pi}{2}(\nu-1)\right)D_{\nu^*}((1+i)\tau))$$
$$+(\exp i\pi\nu)D_\nu((1-i)\tau). \tag{A.6}$$

The Bogoliubov coefficients can be read off from this expression; we find:

$$\beta = \exp(i\pi\nu) = \exp\left(-\frac{\pi}{2}(\lambda + i)\right); \ \alpha = \frac{\sqrt{2\pi}}{\Gamma(\frac{1}{2}(1-i\lambda))}\exp\left(-\frac{\pi}{4}(\lambda - i)\right). \tag{A.7}$$

Note that

$$|\beta|^2 = \exp(-\pi\lambda)$$
$$|\alpha|^2 = \frac{2\pi\exp(-\pi\lambda/2)}{\Gamma(\frac{1}{2}(1-i\lambda))\Gamma(\frac{1}{2}(i+i\lambda))} = 2\exp\left(-\frac{\pi\lambda}{2}\right)\left(\cosh\frac{\pi\lambda}{2}\right) = 1+\exp(-\pi\lambda); \tag{A.8}$$

clearly, $|\alpha|^2 - |\beta|^2 = 1$ as it should.

References

Banks T 1985 *Nucl. Phys.* **B245** 232
Berestetskii V M, Lifshitz E M and Pitaevskii L P 1979 *Quantum electrodynamics* (Vol. 4 of Course of Theoretical Physics, Pergamon), Section 129
Birrel M D and Davies P C W 1982 *Quantum fields in curved space* (Cambridge: University Press)
Capri A Z and Roy S M 1991 preprint Albreta-Thy- 12–91
Chitre D M and Hartle J B 1977 *Phys. Rev.* **D12** 368
Christensen S M and Duff M J 1978 *Nucl. Phys.* **B146** 11
Davies P C W 1975 *J. Phys.* **A8** 365
Duff M J 1981 in *Quantum gravity 2* (eds) C J Isham, R Penrose and D W Sciama (Oxford: University Press)
Feynman R P and Hibbs A R 1965 *Quantum mechanics and path integrals* (New York: McGraw-Hill)
Ford L H 1982 *Ann. Phys.* **144** 238
Fulling S 1973 *Phys. Rev.* **D7** 2870
Fulling S 1989 *Aspects of quantum field theory in curved spacetime* (Cambridge: University Press)
Gibbons G W and Hawking S W 1976 *Phys. Rev.* **D13** 2188
Ginzburg V L 1987 *Issues in intense-field quantum electrodynamics of strong fields* (Berlin: Springer–Verlag)
Gradshteyn I S and Ryzhik I M 1965 *Table of integrals, series and products* (New York: Academic Press)
Greiner W, Müller B and Rafelski J 1985 *Quantum electrodynamics of strong fields* (Berlin: Springer–Verlag)
Hartle J B 1986 *in Gravitation in astrophysics* (eds) J B Hartle and B Carter (New York: Plenum)
Heisenberg W and Euler H 1936 *Z. Phys.* **98** 714
Landau L D and Lifshitz E M 1973 *Quantum mechanics* (Pergamon) p. 184
Lee T D 1986 *Nucl. Phys.* **B264** 437
Marinov M S and Papov V S 1977 *Fortschr. Phys.* **25** 373
Nenciu G and Scharf G 1978 *Helv. Phys. Acta* **51** 412

Nikishov A I 1970a *Sov. Phys. JETP* **30** 660
Nikishov A I 1970b *Nucl. Phys.* **B21** 346
Padmanabhan T 1989 *Class. Quantum Gravit.* **6** 533
Padmanabhan T 1990 *Phys. Rev. Lett.* **64** 2471
Padmanabhan T 1991a TIFR–TAP preprint
Padmanabhan T 1991b *Pramana – J. Phys.* **36** 253
Parker L 1982 *Fundam. Cosmic Phys.* **7** 201
Popov N S 1972a *Sov. Phys. JETP* **35** 659
Popov N S 1972b *Sov. Phys. JETP* **35** 709
Rivers R J 1987 *Path integral methods in quantum field theory* (Cambridge: University Press)
Schwinger J 1951 *Phys. Rev.* **82** 664
Schwinger J 1954a *Phys. Rev.* **93** 615
Schwinger J 1954b *Phys. Rev.* **94** 1362
Seip H P 1982 *Helv. Phys. Acta* **55** 1
Shulman L S 1981 *Techniques and applications of path integration* (New York: Wiley)
Singh T P and Padmanabhan T 1989 *Ann. Phys.* **198** 296
Unruh W 1976 *Phys. Rev.* **D14** 870