BLACK HOLES

AND

QUANTUM COSMOLOGY

Felicity Mellor

A thesis submitted to the University of Newcastle–upon–Tyne
for the Degree of Doctor of Philosophy

Department of Physics

November 1990

ABSTRACT

The work of this thesis falls into two parts: a discussion of decoherence in quantum Kaluza–Klein theories and a study of some of the properties of general black hole metrics in de Sitter spacetime.

Kaluza–Klein theories permit a variety of compactifications and arbitrary scales for the internal space. There must be no quantum interference between these different possibilities. In chapter one it is demonstrated that in the Salam–Sezgin compactification interference between differently scaled interenal spaces is suppressed.

In chapter two new gravitational instantons are presented which are related to charged, rotating black holes with a cosmological constant Λ. These instantons correspond to black holes in de Sitter space with identical Hawking temperatures. Their action contributes terms of order $\Lambda^{-1/2}$ to path integrals with quantum wormholes.

The metrics of these general de Sitter black holes show that the spacetimes have wormholes connecting different asymptotic regions. In chapter three the theory of black hole perturbations is extended to these metrics. It appears that the holes are stable even at the Cauchy horizon. This implies that cosmic censorship is violated. The stability of the spacetimes also implies the existence of a cosmological no hair theorem.

The work presented in this thesis is based on the following papers:

1. *Black Holes and Quantum Wormholes'* with Ian Moss,
   Phys. Lett. **B222** (1989) 361.
2. *Black Holes and Gravitational Instantons* with Ian Moss,
   Class. Quantum Grav. **6** (1989) 1379.
3. *Stability of Black Holes in De Sitter Space* with Ian Moss,
   Phys. Rev. **41** (1990) 403.
4. *Decoherence in Kaluza–Klein Theories*, To be published in Nucl. Phys. **B**.

ACKNOWLEDGEMENTS

I would like to thank all the members of Newcastle Physics Department who have helped create a friendly and stimulating atmosphere in which to work. In particular, thanks to my supervisor, Ian Moss, whose suggestions and guidance have been invaluable.

This work was funded by the Science and Engineering Research Council.

CONTENTS

You see things and you say 'Why?' But I dream things that never were; and I say 'Why not?'

George Bernard Shaw

Back to Methuselah

INTRODUCTION

Prior to the advent of the general theory of relativity in 1915, the consensus view of the scientific community was that the universe was static. As a theory of gravitational dynamics, the general theory provided a new theoretical framework for cosmology – one in which, in its original formulation, the universe was not static. However, relativity was not immediately accepted by the community and it was not until there was observational evidence of the bending of light in a gravitational field that it attracted serious interest. Traditionally held views are, anyhow, a strong force in the shaping of a new theory. Rather than depart from the consensus cosmology, Einstein at first doctored his theory by introducing a repulsive cosmological term to ensure a static universe.

As the general theory matured to become the foundation of twentieth century theoretical cosmology, its artificial manipulation to comply with an ancient aesthetic requirement became less tenable. The observational results of Hubble showed a linear relationship between the radial velocities of Cepheid variables, as measured by the redshifts of their spectral lines, and their distances. This was in compliance with an expanding universe. Such a universe could be modelled through the assumption that on a large scale the universe is spatially homogeneous and isotropic. This appears to be a gross simplification – very large scale structures are known to exist, some spanning 1% of the observable universe. Science, however, operates entirely through simplification. From the available data, attention is focused on a few particular features which are considered to be important, often for quite subjective reasons.

Cosmology, by definition concerned with the spatially and temporally most distant features of the universe, suffers from a great paucity of observational data and is quite beyond the reach of controlled experiment. Most observations are highly model dependent and heavily reliant on other observations. A prime example of this is the hierarchical procedure used in ascertaining extragalactic distances. The difficulties involved in making cosmological observations means that there are only three 'facts' upon which to construct a theory of cosmological evolution:

1. The universe has structure. i.e. there are galaxies and clusters of galaxies.

2. There is a linear relationship between the flux received from the brightest galaxies and their redshift.
3. There is a background of microwave radiation whose spectrum corresponds with great accuracy to isotropic, blackbody radiation at a temperature of 2.7K.

Although the universe can be argued to have structure at very large scales, this point is, to some extent, open to interpretation - it is dependent on how galaxy clusters are counted and on what constitutes structure. Neglecting this possibility then, the three basic observations of cosmology are all consistent with a cosmological principle assuming homogeneity and isotropy. Given the small data set available, cosmologists are in the unusual position among scientists of never having had to debate what data are of particular importance. All independent data become important by virtue of their being available. These data imply the cosmological principle, which in addition has the attraction of greatly constraining the geometric properties of the universe and therefore providing a well–defined and workable starting point for the the construction of a cosmological theory. It is not surprising then that the cosmological principle has been adopted as the underlying principle in modern day cosmology, despite its blatant simplicity.

The cosmological principle, together with the hot big bang model of cosmic nucleosynthesis, has successfully defined a structure within which cosmology has been able to develop both theoretically and observationally. In Kuhnian language the big bang defines a paradigm within which normal science can be pursued. The paradigm enables results to be inferred through the interpretation of observations in the light of the theory and provides a context which directs theoretical progress. The model highlights particular observations as leading to inconsistencies and requiring attention which will probably result in a refinement of the theory.

The most serious problems with the standard scenario have been addressed by inflationary models [1]. These models come in a variety of forms: old, new, chaotic, extended, stochastic and various combinations of these adjectives. They all have in common a de Sitter phase of exponential expansion in the early universe due to the presence of a cosmological constant. The main problems these models seek to explain are:

1. **The horizon problem** - the particle horizons of different parts of the last

scattering surface do not overlap, implying that the isotropy observed today in the microwave background could not have evolved through any physical process but must be assumed as an initial condition. Inflation solves this by altering the structure of the past light cones and thus of the particle horizons.

2. **The flatness problem** - observation sets limits on the density parameter close to the critical value of unity needed for a flat universe. i.e. today, $0.2 < \Omega < 4$. This means that at the Planck time, when the universe emerged from the quatum regime, the universe must have been almost exactly but, unless $\Omega = 1$ today, not actually flat. If the initial conditions of the universe are set at this scale, this requires unnatural fine–tuning. This fine–tuning problem is removed if a sufficiently long period of inflation is introduced.
3. **The formation of structure problem** - observations show galaxies to display structure at a scale comparable to that of the horizon at the time of equal matter and radiation. After this time gravity alone is too weak to build up structures on this scale, but if the structure formed before this it must have reached beyond the particle horizons and thus have evolved acausally. Again, inflation solves this problem by altering the structure of the particle horizons.
4. **The monopole problem** - contrary to observation, the hot big bang predicts a monopole density which would dominate the universe. However, if inflation were to occur shortly after monopole production it would exponentially dilute the rest energy density of monopoles.

Although the original inflationary scenario had some degree of success in solving the major problems of the standard model, it was necessary to assume some flatness at the onset of inflation and the cosmological perturbations predicted exceeded the structure observed. A number of amended versions of inflation have been propsed. However, there is still the problem of finding natural mechanisms by which to realise inflation and by which to 'gracefully exit' from the phase of exponential expansion. Thus many problems remain outstanding in the hot big bang model. Given that there is so little independent data put into the theory, a satisfactory solution to these problems could well require the abandonment of the model altogether, rather than its refinement as in the inflationary approach.

The theory of cosmological evolution might appear vulnerable due to the difficulties involved in obtaining observational data and the absence of controlled experiment. However, results can be obtained within the framework of the model, and technological developments such as the Hubble telescope (had it worked) and the construction of gravitational wave detectors should allow further progress to take place. By contrast, a theory of cosmological origins is an exercise in mathematical creativity at present far removed even from theory dependent observation. The most substantial obstacle to the formulation of such a theory is that any description of the initial conditions of the universe requires a quantum theory of gravity.

According to the big bang model, in its early stages the universe would have had a scale less than that of the Planck scale. Thus only a quantum field theory can describe the behaviour of the universe at this time; general relativity as it stands is inadequate and it is necessary to quantize gravity. Unfortunately, no quantum theory of gravity exists yet, despite rigorous pursuit of such a goal by theoretical particle physicists inspired by the success of the standard model of strong, weak and electromagnetic interactions.

The gravitational field is not amenable to quantization by the usual method. With other field theories, the difficulties involved in producing an exact quantum theory can be avoided by treating interactions as perturbations of the free field theory to obtain an expansion for physical quantities in terms of a power series in the coupling constant. This series is divergent, but the divergencies can be dealt with by introducing cut–offs to produce finite answers, and then taking the limit as the cut–off parameters go to infinity whilst adjusting the bare parameters of the theory to retain the appropriate finite answers. This procedure of renormalisation is necessary for any physically viable perturbative quantum field theory. However, a quantum theory of gravity derived from the perturbation of the spacetime metric is non–renormalisable – the divergent terms cannot be cancelled by the bare terms.

General relativity stands out from the theories of other classical fields due to the special role played by the metric – not only does it describe the structure of the background spacetime, but it also represents the dynamics of the theory. The problems that this dual role introduces makes standard results of quantum field theories inappropriate in the case of quantum gravity. For this reason alternative approaches to the quantisation of gravity have been explored. The canonical approach treats gravity in a similar fashion to the treatment employed in the canonical formulation of quantum mechanics. A Schrödinger type equation is obtained which governs the behaviour of the 'wavefunction of the universe'. The problem of finding a theory of the initial conditions is then reduced to the problem of specifying the boundary conditions of the wavefunction. A number of possible boundary conditions have been suggested – two in particular have stimulated interest. The Vilenkin proposal [2,3] describes a universe which is nucleated from nothing by a tunnelling event. The proposal is formulated most naturally within the canonical framework. However, just as in quantum mechanics there is a powerful alternative to the canonical approach in the form of Feynman's sum–over–histories, a similar alternative formulation has been developed in quantum cosmology. In the path integral approach, the wavefunction of the universe is evaluated by integrating the exponential of the action over all four–geometries and matter configurations on a four–manifold. The action used is the Euclidean action obtained by the analytic continuation of Lorentzian metrics to a Euclidean space of signature (+,+,+,+). For large positive action, the path integral is then exponentially damped.

The Euclidean path integral approach suggests a particularly simple boundary condition. This is the boundary condition of Hartle and Hawking [4-6] which can be summed up in the expression "the boundary condition of the universe is that there is no boundary condition"; that is, the class of manifolds over which the path integral contributions are summed is restricted to those manifolds which have a boundary at the slice of spacetime with which we are concerned, but no other boundary. The Hartle–Hawking proposal has proved popular but it is motivated largely by the aesthetic appeal of its simplicity and the pragmatic appeal of its calculability given certain approximations. However, in the absence of a more satisfactorary theory, the Euclidean path integral approach and the Hartle–Hawking boundary condition do at least provide some sort of formalism to provide a direction for future research.

Both the canonical and path integral approaches to quantum gravity are discussed further in §1.1, and the framework of quantum cosmology together with the Hartle–Hawking prescription is assumed in both chapters one and two.

Having assigned a wavefunction to describe the quantum state of the universe, quantum cosmology raises a number of interpretational questions. Among these is why we actually observe a classical spacetime. This question is directly related to the longstanding problem of measurement in quantum mechanics. Here, it is not obvious how a quantum apparatus can render a classical reading – one would expect to observe interference between the different possible readings. The problem can be rephrased as why the density matrix, given by the modulus squared of the wavefunction, should be diagonal in the basis of the observation. One solution [7,8] to the problem rests on the fact that no system is truely isolated; it is always in interaction with an environment consisting of unobserved modes. The density matrix for the complete apparatus–system–environment combination is not diagonal. However, the physical quantity of interest is actually the reduced density matrix describing the apparatus and system alone. The reduced density matrix is obtained by tracing over the unobserved environment modes. This operation serves to diagonalise the density matrix. This lack of interference in measurements is known as environment induced decoherence.

Environment induced decoherence has been used to understand how classical quantities arise out of a quantum universe. Although the universe is, by definition, an isolated system, there will still be, in any realistic measurement, modes which remain unobserved. These modes constitute the environment. An obvious candidate for the environment modes are the inhomogeneous modes which are typically neglected in calculations.

Chapter one of this thesis studies decoherence in the context of quantum cosmology. In particular, it concentrates on a higher–dimensional Kaluza–Klein theory. Kaluza –Klein theories were first introduced [9,10] as one of the earliest attempts to find a unified theory of the forces of nature. At early times, all dimensions are of a similar size allowing a geometric description of forces other than gravity. At later times, the extra dimensions compactify to a very small scale so that only the remaining four dimensions influence us today. However, in many Kaluza–Klein models, the scale of the compactified space is arbitrary. In a quantum formulation there is therefore the possibility of interference between compactified spaces of different scales. The notion of environment induced decoherence provides a mechanism with which to explore this possibility.

Sections §1.1, §1.2 and §1.3 of chapter one explain in greater detail than the sketches offered here, the basic ideas of quantum cosmology, environment induced decoherence and Kaluza–Klein theories. The remainder of the chapter one focuses on a particular Kaluza–Klein model – the six–dimensional Salam–Sezgin model [11]. The Salam–Sezgin action is computed for a manifold corresponding to a Robertson–Walker universe crossed with a perturbed two–sphere. The perturbations are expanded in terms of harmonics on the two–sphere, as are the lapse function of the metric and the scalar field. From the action, the Hamiltonian and momentum constraints are calculated. Using the semiclassical WKB approximation, an expression for the wavefunction is obtained. This gives a density matrix which, when traced over the perturbation modes is shown to be diagonal and described by a selection function dependent on the scale factors of both the four–dimensional space and the two–dimensional space. If the calculation were extended to include inhomogeneous modes in the four–dimensional spacetime, this generalises to a dependence on the scale factor of the internal space alone. Thus it is concluded that interference between the states of the internal space corresponding to different values of the scale factor is suppressed through environment monitoring.

Quantum cosmology is an attempt at a theory which can describe the physics near the initial singularity of the universe. The general theory of relativity also predicts other types of singularity in the universe. These represent points of the spacetime having infinite curvature and are produced by the complete gravitational collapse of a body. The infinite curvature often means that light cannot escape from bodies which have collapsed to this state, and thus they are known as black holes.These form the subject of chapters two and three. The surfaces beyond which light cannot escape have special properties and are called event horizons.

As with the initial singularity of the universe, quantum effects are important near the singularities within black holes. One of the most important applications of quantum mechanics to a gravitational system has been Hawking's prediction of black hole evaporation [12]. In essence this can be pictured as resulting from quantum fluctuations of the vacuum near the black hole horizon. This prediction gives a physical interpretation to earlier results which found that there are quantities associated with black holes which play roles analogous to those played by the temperature, energy and entropy in thermodynamics. Black holes were found to obey four laws analogous to the laws of thermodynamics. The discovery of the radiation of particles from black hole horizons enabled these analogies to be given a physical motivation and the horizons to be assigned a temperature. The Euclidean formalism is again important in that it can give the particle amplitudes in the black hole spacetime.

In chapter two, the Euclidean sections of general black hole metrics are presented. These black hole metrics represent charged, rotating black holes in de Sitter space. De Sitter spacetime has a cosmological constant which, even in the absence of matter, results in the presence of an event horizon. Like the black hole event horizons, this cosmological event horizon has thermodynamic properties and an associated temperature. A series of co–ordinate transformations can be performed on the metric to show that it is regular at the horizons if the black hole and de Sitter horizons are at the same temperature . Then it is possible to analytically continue the metric to imaginary time and thus obtain a regular Euclidean section. The Euclidean section has finite action and represents a gravitational instanton. Gravitational instantons dominate the sums over positive definite metrics in the Euclidean approach to quantum gravity and are therefore of importance in calculating transition amplitudes. Some of the uses of insantons from both gravitational and other field theories are described in §2.1. Attention is paid, in particular, to those gravitational instantons which involve a non–zero cosmological constant. These instantons have been used by other authors to compute, for instance, the probability of a phase transition occurring to take the universe from a state of false vacuum to one of true vacuum [13-15], to describe the Vilenkin tunnelling from nothing boundary condition and, most recently to provide a picture of transition from a Euclidean universe to a Lorentzian one [16-18].

If there is a magnetic charge on the black hole, then the instantons can be used to evaluate the pair production of monopoles in the de Sitter stage of the early universe. A further application of these instantons concerns quantum wormholes. Recent calculations [20] have offered an explanation of why the cosmological constant is zero by assuming that the universe is joined to itself and to other universes by wormholes or 'spacetime bridges'. These are solutions to the Einstein equations which connect different regions. If they are taken into account in the path integral, then it is found that the fundamental parameters of the theory are no longer fixed, but have a probability distribution . The distribution is peaked around zero cosmological constant. The instanton action plays a large role in determining this distribution. However, it was assumed in the original calculation that the action could be expanded in integral powers of the cosmological constant. The action of the instantons presented here shows that this expansion is not valid – non–integral powers should also enter. This does not alter the leading term of the action and so the mechanism for a vanishing cosmological constant is not affected. But the 'big fix', whereby the higher order terms fix the other parameters of the theory, such as coupling constants, will be affected.

The work recently carried out on transitions from Euclidean to Lorentzian universes [16-18] proposes to include complex metrics in the contour of the path integral. The new instantons presented here confirm that inclusion of complex metrics can be attractive. It allows rotating black hole instantons to contribute to the path integral.

The black hole metrics discussed in §2.2 have interesting global properties. The spacetime contains a timelike singularity and three horizons: the de Sitter horizon, the outer black hole event horizon and the inner black hole horizon or Cauchy horizon. The Cauchy horizon represents the limit of predictability from some surface of initial data. Beyond this horizon causality is violated and the singularity is no longer hidden from view by the black hole. If an observer could traverse the Cauchy horizon, they would be able to see the naked singularity and this would contravene the cosmic censorship conjecture which states that all singularities in the universe, with perhaps the exception of the initial singularity, are hidden. To a large extent, this conjecture motivates the study of black holes.

The features of a Cauchy horizon and a timelike singularity in the metric of §2.2 are also found in the Reissner–Nordström metric – the metric of a charged black hole in flat spacetime. In both cases the Cauchy horizon serves to join one asymptotic region of the universe to another: it acts as a wormhole. In principle, an observer could cross the Cauchy horizon, travel past the naked singularity without falling into it and emerge in another universe. Such a journey would introduce many conceptual difficulties and it is therefore of interest to discover whether such a journey is actually possible. This has been done in the case of the Reissner–Nordström black hole by considering the behaviour of perturbations near the Cauchy horizon using the theory of potential scattering [20,21]. This shows that the horizon is unstable since the perturbation is unbounded. This can be explained by the fact that the observer approaching the horizon sees the entire future of the universe outside in only a finite time interval. The radiation coming from outside becomes infinitely blue–shifted, compressed to an infinite energy density on the horizon. In a closed universe represented by de Sitter space this argument no longer applies. The global structure of the spacetime is altered by the replacement of null infinity with the de Sitter horizon. Since the Cauchy horizon is the continuation of this surface within the black hole, the observer at the Cauchy horizon now sees only the finite number of events which lie within the de Sitter horizon. It is possible, then, that the Cauchy horizon of a charged black hole in de Sitter space could be stable.

This possibility is explored in chapter three, again by employing the theory of potential scattering to study the growth of perturbations. The equations pertinent to the study are derived in §3.3. Their implications are then examined for both empty de Sitter space and a black hole–de Sitter space in sections §3.4 and §3.5. Although empty de Sitter space has no Cauchy horizon, it is still of interest to study the behaviour of perturbations here. The stability of de Sitter space is an important requirement of the inflationary scenario in which the universe must approach de Sitter space if it is to undergo the necessary period of exponential expansion. This requirement has become known as the cosmological no hair theorem. It is similar to the no hair theorems of black holes which show that all possible outcomes of complete gravitational collapse in flat space are described by the known Kerr–Newman metrics describing spherically symmetric, rotating, charged black holes. Thus the shape of the collapsing body has no bearing on the final outcome of collapse and all black holes can be parametrised by their mass, charge and angular momentum alone. Unlike the cosmological no hair theorems, the black hole theorems have been proved rigorously. No hair theorems are discussed further in §3.2.

In §3.4 it is shown that perturbations in de Sitter space remain finite. However, it can be argued that the expansion of the universe has the effect of causing the wavelength of the perturbations to grow so that it is eventually greater than the horizon size. In this case, the spacetime can be said to be stable.

The considerations of perturbations incident upon the Cauchy horizon show that this is also stable. In previous studies [20,21], it has been shown that the poles in the reciprocal transmission coefficient for perturbation modes in the Reissner–Nordström spacetime render the Cauchy horizon unstable. However, this argument fails to allow for the possibility that the perturbations entering the black hole from outside may have a zero in their amplitude which cancels the relevant pole. Unlikely as it may seem, §3.5 shows that this is exactly what happens for a black hole in de Sitter space. There remains the possibility that there may be poles in the transmission coefficients of the modes exterior to the hole. These poles are the quasi–normal modes of the black hole. Numerical evidence is presented here that shows these modes do not hinder the stability of the Cauchy horizon for black holes whose charge and mass are equal or lie within a narrow range of each other.

The stability of the Cauchy horizon of these black holes implies that it is possible to see naked singularities. That it is not possible to see one is one of the basic assumptions of classical relativity. The results of chapter three suggest that classical relativity as it stands is inadequate. This is similar to the conclusions drawn from studies of the initial singularity where quantum cosmology must be introduced to arrive at any description of the physics there.

The analysis of chapter three was carried out in a universe closed due to the presence of a cosmological constant. Although such a spacetime is of importance due to interest in the inflationary scenario, it would perhaps be of even greater interest to carry out a similar stability analysis in a universe closed by more natural means; for instance, in a matter closed universe. This could be a topic for further research.

Notation

In this thesis, the metric signature (−,+,+,+) is adopted unless otherwise stated.

In chapter one, latin letters are used to label the internal space, greek letters from the middle of the alphabet label the four–dimensional universe, whilst greek letters from the beginning of the alphabet label the complete (4+d)–spacetime. A vertical bar indicates covariant differentiation with respect to the unperturbed two–sphere.

CHAPTER ONE

DECOHERENCE IN QUANTUM KALUZA–KLEIN THEORIES

1.1 Quantum cosmology

Any reasonable theory of a physical system consists of a collection of dynamical laws together with suitable initial conditions. If the system in question is the whole universe, the derivation of a theory of initial conditions is problematic. At early times the small scale of the early universe necessitates the construction of a quantum theory of gravity; the boundary conditions of which would provide the required initial conditions. However, no consistent theory of gravity exists as yet. Quantum cosmology sidesteps many of the technical difficulties involved in constructing a quantum field theory for general relativity and, using a dynamical law analogous to the Schrödinger equation,proceeds directly to the issue of the boundary conditions.

The construction of the appropriate dynamics for quantum cosmology is most clearly seen in the canonical Hamiltonian formulation. As in the Hamiltonian approach to any field theory, this requires splitting the spacetime into space and time. The physical variable of any theory hoping to produce Einstein's general theory of relativity in the classical limit must be the four–metric $g_{\mu\nu}$ of spacetime. The co–ordinate independence of the classical gravitational action suggests that $g_{\mu\nu}$ may be decomposed into three–surfaces of constant co–ordinate, t. t is called 'time', but this is no more than a label. Physical time is signified by the correlations of clocks with the system, which requires a knowledge of the spacetime metric. Since for gravity this is the variable of the theory, no such knowledge is available and so 'time' has no physical meaning. The foliation of spacetime implies that it can be described entirely in terms of the three–metric, h_{ij}, on the spacelike hypersurfaces and the way in which these hypersurfaces fit together. The latter can be given by the lapse and shift functions of the Arnowitt, Deser and Misner (ADM) formalism [22]. The lapse in proper time between a hypersurface at t and another at $t + dt$ is given by $N(x)\,dt$ where $N(x)$ is the lapse function. The shift function, N^i, gives the correspondence between points in the two hypersurfaces. The point $(x^i + dx^i + N^i dt)$ in the lower hypersurface (that at t) corresponds to the point $(x^i + dx^i)$ in the upper hypersurface. Given the lapse and shift functions and the three–metric on the hypersurface, the four–metric can be written in the (3+1) form

$$ds^2 = -(Ndt)^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \; . \tag{1.1.1}$$

If a theory of quantum gravity is to be developed in the language of the ADM formalism, then it is necessary to express the concept of curvature in (3+1) form. This leads to the distinction between intrinsic and extrinsic curvature. The intrinsic curvature of a three–surface, ${}^3R$, is the usual concept of curvature encountered in differential geometry. The extrinsic curvature, K, however, has no meaning for a three–surface in itself. Rather, it is a measure of curvature relative to an embedding and gives the relative deformation of a figure lying in the surface as each point is carried forward a unit interval of proper time normal to the hypersurface. The extrinsic curvature is equivalent to the second differential form of differential geometry and is therefore given by the expression

$$K_{ij} = \frac{1}{2N}\left[N_{i;j} + N_{j;i} - \frac{\partial h_{ij}}{\partial t}\right] , \tag{1.1.2}$$

where a semicolon represents covariant differentiation with respect to the three–geometry h_{ij}.

The curvature of the embedding four–manifold can be related to the intrinsic and extrinsic curvatures of the spacelike hypersurfaces resulting from foliation through the Gauss–Codazzi equations. This enables the Hilbert Lagrangian density

$$16\pi G\mathcal{L} = \sqrt{-{}^4g}\,{}^4R \tag{1.1.3}$$

to be rewritten in the form

$$\begin{aligned} 16\pi G\mathcal{L} =& N h^{1/2}\left({}^3R + K_{ij}K^{ij} - (\operatorname{Tr} K)^2\right) \\ & - 2\frac{\partial}{\partial t}\left(h^{1/2}K\right) + 2\frac{\partial}{\partial x^i}\left(h^{1/2}\left(KN^i - h^{ij}N_{;j}\right)\right) \; . \end{aligned} \tag{1.1.4}$$

When the Lagrangian is integrated to obtain the classical action, the last two terms, which are total derivatives, integrate out to give surface terms. Since these have no effect on the equations of motion they can be dropped.

Expression (1.1.4) for the Lagrangian density enables one to implement the canonical formalism in the usual way. The canonical momentum conjugate to h_{ij} is

$$\pi^{ij} = \frac{\partial \mathcal{L}}{\partial \dot{h}_{ij}} = \sqrt{h}\left(K^{ij} - Kh^{ij}\right) . \tag{1.1.5}$$

Since no time derivatives of the lapse and shift functions appear in the Lagrangian, their conjugate momenta both vanish. i.e.

$$\pi^{i} = \frac{\partial \mathcal{L}}{\partial \dot{N}_i} = 0 \, ; \quad \pi = \frac{\partial \mathcal{L}}{\partial \dot{N}} = 0 \, . \tag{1.1.6}$$

This implies that neither the lapse nor shift functions should be viewed as a dynamical variable of the theory. They are Lagrange multipliers which enforce constraints upon the Hamiltonian system.

The action can be written in the form

$$S = \int d^4x \left[\pi^{ij}\dot{h}_{ij} - N\mathcal{H} - N_i\mathcal{H}^i\right] , \tag{1.1.7}$$

where $\mathcal{H}^i$ is the 'supermomentum' given by

$$\mathcal{H}^i = -2\pi^{ij}_{\;;j} \tag{1.1.8}$$

and $\mathcal{H}$ is the 'superhamiltonian' given by

$$\mathcal{H} = 16\pi G.G_{ijkl}\pi^{ij}\pi^{kl} - \frac{1}{16\pi G}\sqrt{h}\,{}^3R \tag{1.1.9}$$

with

$$G_{ijkl} = \tfrac{1}{2}h^{-1/2}\left(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}\right) . \tag{1.1.10}$$

Variation of the action with respect to the Lagrange multipliers N and N^i yields two constraint equations: the Hamiltonian constraint

$$\mathcal{H} = 0 \, ; \tag{1.1.11}$$

and the momentum constraint

$$\mathcal{H}^i = 0 \, . \tag{1.1.12}$$

A procedure for the quantisation of the theory can now be adopted which is analogous to that employed in the derivation of the Schrödinger equation in quantum mechanics. This simply involves the operator replacement

$$\pi^{ij} \rightarrow -i\frac{\delta}{\delta h_{ij}} \tag{1.1.13}$$

in the Hamiltonian. This introduces some factor ordering ambiguities, but these can largely be ignored since they have little effect on the results of calculations. Usually the ordering of operators is chosen as

$$\pi^{ij}\pi_{ij} \rightarrow \frac{1}{\sqrt{G}}\frac{\delta}{\delta h_{ij}}\sqrt{G}\, G^{ijkl}\frac{\delta}{\delta h_{kl}} \,. \tag{1.1.14}$$

With the substitution (1.1.13), the constraints (1.1.8) and (1.1.9) become operator restrictions on the wavefunction Ψ. The Hamiltonian constraint yields the Wheeler–De Witt equation

$$\mathcal{H}\Psi = \left(-16\pi G\,\frac{1}{\sqrt{G}}\frac{\delta}{\delta h_{ij}}\sqrt{G}\, G^{ijkl}\frac{\delta}{\delta h_{kl}} - \frac{1}{16\pi G}\sqrt{h}\,{}^3R\right)\Psi = 0\,. \tag{1.1.15}$$

The momentum constraint,

$$\mathcal{H}^i\Psi = \left(\frac{\delta\Psi}{\delta h_{ij}}\right)_{;j} = 0\,, \tag{1.1.16}$$

implies invariance under three–dimensional diffeomorphisms. That is, although the wavefunction depends on the three–geometry, it is not dependent on the particular three–metric h_{ij}. The situation is similar to that which arises in the Hamiltonian formulation of Maxwell's equations. There, the constrained nature of the Hamiltonian formulation indicates that there is a gauge freedom in the choice of variable A^μ. Similarly, the fact that the Einstein equations produce a constrained Hamiltonian system suggests that there is a gauge arbitrariness in the choice of field h_{ij}. The correct configuration space for general relativity is therefore the set of equivalence classes of three–metrics. This set is known as 'superspace' and has metric G_{ijkl}.

In the presence of matter, the wavefunction is a functional of all three–metrics, h_{ij}, and all matter fields, ϕ, defined on the three–surface. i.e.

$$\Psi = \Psi[h_{ij}, \phi]\,. \tag{1.1.17}$$

As has been seen, the wavefunction obeys a system of zero energy Schrödinger equations defined on superspace. The matter fields result in a further potential term in (1.1.15).

Unfortunately, the infinite–dimensional nature of superspace makes full calculations impracticable. Instead, a finite–dimensional approximation is often adopted , known as 'minisuperspace'. Usually this approximation consists in a reduction of the number of degrees of freedom through a consideration of isotropic and homogeneous metrics only. It is then possible to solve the Wheeler–De Witt equations.

The solution Ψ to the Wheeler–De Witt equation is called the wavefunction of the universe and it represents possible quantum states of the universe. The problem of finding a law for the initial conditions is equivalent to that of specifying those boundary conditions which select from the possible solutions of the Wheeler–De Witt equation that wavefunction which corresponds to our universe. A number of proposals have been made for the boundary condition of the universe. In particular, interest has concentrated on the 'no–boundary' propsal of Hartle and Hawking [4-6] and the 'spontateous creation from nothing' proposal of Vilenkin [2,3].

The Hartle–Hawking proposal is formulated in terms of a functional approach to the quantization of general relativity. Just as Feynman's sum–over–histories provides an alternative to the Hamiltonian formulation of quantum mechanics, a similar approach can be adopted as an alternative to the Hamiltonian formulation of general relativity. In this case, the histories of concern are the four–geometries $g_{\mu\nu}$ and matter field configurations ϕ on a four–manifold M. The wavefunction can be expressed as the path integral

$$\Psi[h_{ij}, \phi] = \sum_M \int_C \mathcal{D}g_{\mu\nu}\mathcal{D}\phi \, e^{-S_E[g_{\mu\nu},\phi]} \,. \tag{1.1.18}$$

S_E is the Euclidean action obtained by setting N negative imaginary. In this formalism it is necessary to include the surface term

$$-\frac{1}{16\pi G}\int_{\partial M} 2K\sqrt{h}d^3x \tag{1.1.19}$$

in the Euclidean action. Although such terms will vanish for metrics approaching zero fast enough at infinity, the functional integral must be performed over *all* metrics meeting adequate boundary conditions in Euclidean spacetime. In general then, these terms will contribute.

It is possible to recover the Wheeler–De Witt equation from the functional formalism. As explained above, the effect of a variation in the lapse function N is to push the hypersurface forwards or backwards in time. However, the wavefunction should be independent of time - it is a functional only of the intrinsic geometry of the hypersurface and the fields thereon. This implies that

$$0 = \sum_M \int_C \mathcal{D}g_{\mu\nu}\mathcal{D}\phi \, \frac{\delta S_E}{\delta N} e^{-S_E[g_{\mu\nu},\phi]} \, . \tag{1.1.20}$$

(1.1.20) is essentially the Wheeler–De Witt equation since

$$\frac{\delta S_E}{\delta N} = \mathcal{H} \, . \tag{1.1.21}$$

Likewise, variation of the wavefunction with respect to the shift function, N^i, yields the condition

$$0 = \sum_M \int_C \mathcal{D}g_{\mu\nu}\mathcal{D}\phi \, \frac{\delta S_E}{\delta N_i} e^{-S_E[g_{\mu\nu},\phi]} \, , \tag{1.1.22}$$

which implies invariance under diffeomorphisms as in the Hamiltonian formulation.

The class of manifolds included in the sum in (1.1.18) constitutes a particular choice of boundary conditions. In the Hartle–Hawking proposal the suggestion is to include all compact manifolds which have one boundary at ∂M but no other boundary. This proposal is the analogue for closed cosmologies of the ground state in quantum mechanics. It represents the state of minimum excitation. It seems reasonable to assume that the early universe was in a state of low excitation and so the Hartle–Hawking proposal is an appealing one.

The path integral can also be used to define amplitudes for changes in topology. This gives rise to topologically exotic features such as wormholes and spacetime foam which is discussed further in chapter two. The inclusion of manifolds with wormholes in the path integral might play an important role in determining the coupling constants of the low energy theory of elementary particles [19]. This provides a strong physical motivation for the adoption of the Hartle–Hawking prescription. A frequently quoted attraction of the Hartle–Hawking proposal is that it is a workable proposal. Using this boundary condition it is possible to use a WKB expansion to solve the Wheeler–De Witt equation in a suitable minisuperspace approximation. In general, the solution may be oscillatory or exponential, depending on the boundary conditions. The regions of minisuperspace giving oscillatory solutions are thought to correspond to classically allowed regions, whilst the exponential regions are classically forbidden. In the oscillatory region the Wheeler–De Witt equation can be solved using the semi–classical approximation of the WKB expansion. Thus

$$\Psi = \sum_k C_k e^{iS_k} , \tag{1.1.23}$$

where S_k is a rapidly varying phase and C_k is a slowly varying amplitude. Then S_k satisfies the Hamilton–Jacobi equations which correspond to the classical solutions of the field equations.For the Hartle–Hawking proposal, the wavefunction has two WKB components,

$$\Psi_{HH} = Ce^{iS} + C^* e^{-iS} , \tag{1.1.24}$$

representing a superposition of expanding and collapsing universe.

The wavefunction studied in §1.5 is assumed to be initially in the Hartle–Hawking state

1.2 Environment Induced Decoherence

Quantum cosmology aims to describe the whole universe using the rules of quantum theory. This universality of the theory would require that all phenomena be explicable within a quantum framework, including the highly classical behaviour of the macroscopic world. Classical theories should be entirely contained within quantum ones. In the classical description of a system, one can have definite knowledge of the state of that system to a chosen degree of accuracy. Given this knowledge, together with a suitable dynamical law, one may then calculate unambiguously the state of the system at any instant thereafter. In quantum mechanics, however, such an unambiguous calculation is not possible. Not only is quantum theory probabilistic, but the quantum state of a system is described by amplitudes subject to the principle of superposition. Thus, by contrast to classical probabilistic theories such as thermodynamics where probabilities can be summed, in quantum mechanics it is the probability amplitudes which are summed, giving rise to interference terms between the different states.

In discussions of the quantum theory of measurement this becomes problematic. Here, one needs to understand how an observation on a quantum apparatus might render a classical reading. In particular, why does one not observe interference between different possible readings. Clearly, before one can reach any such understanding, one must examine precisely what is involved in the act of measurement. Any measurement consists of a system, $\mathcal{S}$, which is to be observed, and a measuring apparatus, $\mathcal{A}$. Following von Neumann's analysis [25], the measurement may be divided into two distinct stages:

1) In the first stage, the system and apparatus, initially isolated, are brought into contact. There proceeds an interaction which obeys the usual quantum mechanical laws. That is, the total wavefunction of the combined system and apparatus obeys Schrödinger's equation. The interaction disturbs the system $\mathcal{S}$ little, but enables the state of the apparatus to become correlated with that of $\mathcal{S}$. However, this is not sufficient for an observation. Given a state of $\mathcal{S}$, the apparatus $\mathcal{A}$ can now be said to be in a certain 'relative' state [26], but one is none the wiser as to what the state of $\mathcal{S}$ is. Moreover, at this stage the system–apparatus density matrix contains non–zero off–diagonal terms. These represent interference between the different possible outcomes of the measurement. For a classical observation, these are precisely the terms which must be eliminated. This is not possible under unitary Schrödinger evolution. Thus a second stage to the act of measurement must be invoked.

2) In the second stage a discontinuous, non–unitary transformation occurs which enables the density matrix to become diagonal. Thus there is now no interference and the apparatus may record a classical reading. This does not mean that all probabilistic elements of the quantum system have been removed. The density matrix, which encompasses all information that may be known about a given system, tells only the probability of a certain state being recorded. However, probabilities in this sense also occur in classical statistics and are not in themselves problematic. The problem lies in the mechanism whereby this second stage comes about. The Copenhagen Interpretation suggests a mechanism in which the wavefunction is collapsed

through the fundamental role played by the observer. In this way the second stage of measurement is effectively relegated to beyond the quantum domain. This seems unsatisfactory, particularly in the context of quantum cosmology. We seek a universal quantum theory which will describe all aspects of measurement and which will encompass observers themselves. The alternative theory of Everett [26] denied the fundamental division of the universe into observer and observed, allowing for a quantum description of all stages of measurement. The proposal of many worlds accepts the probalistic nature of the diagonal density matrix by assuming the existence of an ensemble of universes in which the different possible outcomes of a measurement are realised with the appropriate probabilities. The question remains, however, as to how the non-unitary transformation occured. How is the density matrix diagonalised?

This is essentially the problem with the theory of measurement. A solution has been proposed by a number of authors, particularly by Zurek [7], [8]. It is this approach which will be adopted in this chapter. However, before outlining Zurek's analysis, it will be beneficial to look at the first stage of measurement in a little more detail.

Suppose the observed system, $\mathcal{S}$, is described by the set of state vectors $\{|s_n\rangle\}$, and the measuring apparatus, $\mathcal{A}$, by the state vectors $\{|a_n\rangle\}$. Further, suppose that before $\mathcal{A}$ and $\mathcal{S}$ are coupled, $\mathcal{A}$ is in the state $|a_0\rangle$ and $\mathcal{S}$ is in a linear combination of states given by $\sum_n c_n|s_n\rangle$. Initially, there being no interaction Hamiltonian, the total wavefunction of the $\mathcal{A}-\mathcal{S}$ system is given by the product

$$|\Psi_i\rangle = |a_0\rangle \sum_n c_n|s_n\rangle = \sum_n c_n|a_0\rangle|s_n\rangle \ . \tag{1.2.1}$$

The nature of a measurement is such that the state of the system $\mathcal{S}$ should remain unaltered after interaction with the apparatus $\mathcal{A}$. The only effect of the measurement is to correlate $\mathcal{A}$ with $\mathcal{S}$. Thus the final wavefunction may be written as

$$|\Psi_f\rangle = \sum_n c_n|a_n\rangle|s_n\rangle \ . \tag{1.2.2}$$

To see more clearly the implications of (1.2.2) one must turn to the density matrix. Here, this takes the form

$$\rho = |\Psi_f\rangle\langle\Psi_f| = \sum_{n,m} c_n c_m^*|a_n\rangle|s_n\rangle\langle s_m|\langle a_m| \ . \tag{1.2.3}$$

This is the density matrix of a system in a pure state. That is, the system is described by a linear superposition of states and thus the density matrix contains off–diagonal terms representing interference between these states. It is of little surprise that this is the case, since the unitary evolution of the Schrödinger equation cannot disturb the superpositional nature of the system. However, as stated above, we seek a diagonal matrix as the result of measurement. This means that the system must be in a mixed state. A mixed state may be thought of as a virtual ensemble of identical systems in different pure states, each occurring with a relative weight $p_n = |c_n|^2$ [27]. The density matrix is then of the required form

$$\rho_{\text{mixed}} = \sum_n |c_n|^2 |a_n\rangle|s_n\rangle\langle s_n|\langle a_n| \, . \tag{1.2.4}$$

Thus the problem of measurement may be reformulated as the need for some sort of transformation from a pure state density matrix to a mixed state density matrix.

(1.2.2) reveals a further limitation in the first stage of measurement. The rules of quantum mechanics permit an arbitrary choice of basis. Thus, since the apparatus $\mathcal{A}$ is quantum, one may change to a new orthonormal basis given by

$$|\tilde{a}_m\rangle = \sum_n |a_n\rangle\langle a_n|\tilde{a}_m\rangle \, . \tag{1.2.5}$$

The final wavefunction in (1.2.2) then becomes

$$|\Psi_f\rangle = \sum_n c_n \sum_m |\tilde{a}_m\rangle\langle\tilde{a}_m|a_n\rangle|s_n\rangle = \sum_m \tilde{c}_m|\tilde{a}_m\rangle|\tilde{s}_m\rangle \, . \tag{1.2.6}$$

So, even if one could say that the $\mathcal{A}-\mathcal{S}$ system were in a definite distribution of states, one could not assign any such definite distribution to either the apparatus or the observed system without reference to the other.

Thus, not only does the first stage of measurement fail to eliminate interference terms, it also fails to eliminate the arbitrariness introduced through the freedom of choice of apparatus basis. It would appear that one cannot specify whether the system $\mathcal{S}$ is in one of the states $\{|s_n\rangle\}$ or one of the states $\{|\tilde{s}_m\rangle\}$. This is contrary to actual experience where the observable measured on the apparatus is not an arbitrary choice which may be made some time after the apparatus-system interaction. Rather, apparatuses are regarded as having some unique 'pointer basis' which determines the relative states for the observable to be measured.

The key feature of Zurek's approach to the problems of measurement is that even the total $\mathcal{A} - \mathcal{S}$ system cannot be treated as isolated. For every apparatus and system under observation, there is an environment, $\mathcal{E}$, which is not observed but which must be taken into consideration in any analysis of the act of measurement. The effect is to enable off–diagonal terms in the pure state density matrix to become very small, giving a good approximation to a mixed state density matrix. In this way, the density matrix decoheres.

To see this, suppose that the environment is a quantum system represented by states $\{|e_n\rangle\}$, and that it is initially in state $|e_0\rangle$. Then the initial state of the total $\mathcal{S} - \mathcal{A} - \mathcal{E}$ system is

$$|\Phi_i\rangle = \sum_n c_n |s_n\rangle|a_0\rangle|e_0\rangle. \qquad (1.2.7)$$

As before, the premeasurement interaction correlates the apparatus with the system $\mathcal{S}$. This occurs over a short time only, after which any $\mathcal{A} - \mathcal{S}$ interction is assumed negligible. On a longer time–scale, however, there is a further interaction – that of the apparatus with the environment. This serves to correlate the environment with the apparatus. Thus the final state of the total system is given by

$$|\Phi_f\rangle = \sum_n c_n |s_n\rangle|a_n\rangle|e_n\rangle \ . \qquad (1.2.8)$$

It has been assumed here that the system $\mathcal{S}$ has no interaction with the environment. This ensures that there is no change in the state of $\mathcal{S}$ after premeasurement due to the influence of the environment, so one is able to measure the present state of $\mathcal{S}$ and not some past state.

The expression for the final state once again enables one to write down the density matrix. As before, since only unitary evolution has occurred, the total density matrix will describe a pure state. However, this density matrix refers to the total $\mathcal{A} - \mathcal{S} - \mathcal{E}$ system. This is not what is actually observed. The quantity of interest, therefore, is not the total density matrix, but a reduced density matrix in which the environment states have been traced over.

$$\rho_{\text{reduced}} = Tr_{\mathcal{E}}|\Phi_f\rangle\langle\Phi_f| = \sum_{n,m} c_n {c^*}_m \ \langle e_n|e_m\rangle|s_n\rangle|a_n\rangle\langle a_m|\langle s_m| \ . \qquad (1.2.9)$$

From (1.2.9) it can be seen that the reduced density matrix represents a mixed state. If the inner products $\langle e_n | e_m \rangle$ are negligible for $n \neq m$, then the matrix will be approximately diagonal. Hence the environment causes the density matrix to decohere so that any interference terms in the $\mathcal{S} - \mathcal{A}$ system disappear.

Not only does this continuous measurement of the apparatus by the environment result in decoherence, it also uniquely determines the pointer observable of the apparatus. The information about the state of $\mathcal{S}$ held in the apparatus as a result of correlation during premeasurement will be destoyed during the longer apparatus–environment interaction unless the Hamiltonian of the interaction, $\mathcal{H}_{\mathcal{A}-\mathcal{S}}$, commutes with the observable of the apparatus. Only the eigenbasis of this observable will record nothing but information about the state of $\mathcal{S}$.

In this manner the environment picks out a unique basis for the pointer observable and the eigenvalues of this observable are the elements of a diagonal density matrix. In a sense, the environment acts as a second apparatus which continously monitors the first. However, the introduction of a secondary apparatus alone would not be sufficient to resolve the problems of decoherence and choice of pointer basis. As pointed out by von Neumann [25], additional apparatuses might remove these problems from the consideration of the original apparatus, but they would have to be faced eventually in one of the additional apparatuses. It is only the fact that the environment states go unobserved that endows the system with non-unitary evolution to a mixed state.

At first sight it appears strange that an understanding of quantum measurement, based entirely on the fact that no system–apparatus combination can be thought of as isolated, might be of any relevance to quantum cosmology. Here, surely, is a system which certainly is isolated – by definition there is nothing exterior to a closed universe. Moreover, there is not even a clear distinction between the apparatus and the system under observation. The latter, being the universe itself, must contain the former. However, decoherence, as understood through environment monitoring, has proved a fruitful approach to the study of classical properties arising from a quantum universe. The lack of an obvious candidate for the environment need not present a problem. To some extent, the boundaries of the apparatus, system and environment are arbitrary. For instance, since it is ultimately the apparatus–system combination which is observed, it is necessary only that there be a line dividing them initially, but exactly where that line is drawn is not of importance. Similarly, the concept of the environment contains some arbitrariness. The constraint that it should be comprised of unobserved states still leaves many degrees of freedom which can in fact be ignored since they do not interact with the apparatus. The environment, then, is defined in a somewhat pragmatic fashion, and the immediate environment is distinguished from the remote environment which may be neglected. In quantum cosmology a model can contain many variables which are not of direct interest. Minisuperspace, for instance, concentrates on the homogeneous modes of the fields. It is natural, therefore, to turn to the unobserved inhomogeneous modes to fulfil the role of the environment and to regard the rest of the universe as the system under observation.

This technique has been employed by a number of authors [28-33] examining decoherence in quantum cosmology. In particular, the way in which the scale factor is rendered classical has been studied. The first detailed calculation of this sort was by Kiefer [28], who looked at a Friedmann universe and considered the environment to be the perturbations of the 3-metric and the massive scalar field. The continuous measurement by these higher multipoles was shown to produce perfectly classical behaviour in the scale factor and approximately classical behaviour in the matter field defined by a narrow Gaussian. Decoherence of the scale factor has also been demonstrated by Halliwell [29] in a de Sitter model with the environment as the inhomogeneous modes of a massless scalar field. He also showed that there is no interference between the different components in the WKB approximation corresponding to the expanding and contracting phases of the universe. It is justifiable, therefore, to consider only one such component.

The procedure adopted in the following calculation is similar to that of the above mentioned papers. Decoherence is studied in the context of a six–dimensional Kaluza–Klein theory. However, before describing the particular model of interest, I will outline some of the properties of Kaluza–Klein theories generally.

1.3 Classical Kaluza–Klein Theories

The motivation behind Kaluza–Klein theories lies in the desire to formulate a unified description of the forces of nature. The original theories of Kaluza [9] and Klein [10] sought to provide a geometric formulation of electromagnetism through a theory of Einstein gravity in five dimensions. The theory naturally reduces to Minkowski space with a compact circle S^1, so the ground state metric is

$$g^{(0)}_{\alpha\beta} = \begin{pmatrix} \eta_{\mu\nu} & 0 \\ 0 & g_{55} \end{pmatrix} , \tag{1.3.1}$$

where $\eta_{\mu\nu} = (-1,1,1,1)$ is the metric of Minkowski space and $g_{55} = R^2$ is the metric of the manifold S^1 with R the radius of the circle. From a generalisation of this metric, one can derive a theory of gravity in four–dimensions and a theory of an Abelian gauge field. To see this, the dependence of the metric on the co–ordinate θ parametrizing the extra dimension can be neglected. A general ansatz for the metric is then

$$g_{\alpha\beta} = \begin{pmatrix} g_{\mu\nu}(x) + \alpha^2 A_\mu(x) A_\nu(x) g_{55} & -\alpha A_\mu(x) g_{55} \\ & \\ -\alpha A_\nu(x) g_{55} & g_{55} \end{pmatrix} , \tag{1.3.2}$$

where α is a scale factor chosen as $16\pi G/R$ for convenient renormalization. Under a co–ordinate transformation of the compact manifold co–ordinate,

$$\theta \longrightarrow \theta' = \theta + \alpha\, \epsilon(x) , \tag{1.3.3}$$

the off–diagonal elements of the metric give

$$A_\mu \longrightarrow A'_\mu + \partial_\mu \epsilon . \tag{1.3.4}$$

Thus the transformation of co–ordinates in the extra spatial dimension induces an Abelian gauge transformation on A_μ. In this manner, a candidate theory for electromagnetism arises alongside four–dimensional gravity from a simpler five–dimensional theory.

Moreover, the approach accomodates the quantization of charge of any associated fields. For instance, a five–dimensional massless scalar field can be Fourier expanded on the compact circle:

$$\phi(x,\theta) = \sum_{n=-\infty}^{\infty} \phi^n(x)\, e^{in\theta} . \tag{1.3.5}$$

Then, under the co-ordinate transformation (1.3.3), the field $\phi(x,\theta)$ transforms as

$$\phi^n(x) \longrightarrow e^{in\alpha\epsilon(x)}\phi^n(x)\,. \tag{1.3.6}$$

Thus the Lagrangian of the field will be invariant under the gauge transformation of the vector field A_μ, and a quantized charge, q, can be identified, where

$$q = -n\alpha = -\frac{16\pi G n}{R}\,. \tag{1.3.7}$$

In the case of electromagnetism, where the unit of charge q/n is equal to the electric quantum e, (1.3.7) imposes a constraint on the size of the fifth dimension:

$$R \sim 10\, m_p^{-1}\,. \tag{1.3.8}$$

The small size of the extra dimension enables one to explain the quantization of electric charge without expecting to observe a higher dimension.

Of course, the original Kaluza–Klein theory was conceived at a time when only the electromagnetic and gravitational forces were recognised. Today, if one is to pursue the quest for a unification scheme arising out of a higher dimensional theory, then that theory should also encompass the weak and strong interactions. The compact manifold should provide the internal symmetry for a non-Abelian gauge group. This is possible by looking at isometries of a manifold of dimension greater than one [34]. A straightforward generalization of the five–dimensional case shows that these isometries can generate the Yang–Mills gauge transformation.

However, a serious fault with both the original theory and it's extensions to higher dimensions, is that the compactified metric is not a solution to the classical higher dimensional Einstein equations. It is necessary, therefore, to explicity invoke some mechanism which induces a dynamical compactification. One such mechanism is to introduce additional gauge fields into the theory at the higher dimensional level. The presence of a non–zero energy–momentum tensor in (4+d)–dimensions is sufficient to ensure that the classical solution of the field equations is a product of a four–dimensional spacetime with some compact internal space [35]. It can then be said that spontaneous compactification has occurred. The resulting theory, being dynamical, is more powerful and realistic than those in which one must forego solving

Einstein's equations if compactification is to be fulfilled. However, the introduction of fields other than gravity into the higher dimensional theory is both arbitrary and against the spirit of Kaluza–Klein theories. The arbitrariness is most easily removed by looking to supersymmetry which provides natural candidates for the extra bose fields. Indeed, historically it was found that theories of supergravity were most easily formulated in dimensions greater than four, and it was this that led to a revival of interest in the ideas of Kaluza–Klein theories during the seventies.

One theory which has attracted particular interest in this context is eleven–dimensional supergravity. Not only is an eleven–dimensional manifold the largest that will admit supergravity, but it is also the smallest for a Kaluza–Klein theory producing the gauge group of grand unification [36]. For this reason eleven–dimensional supergravity has stimulated much interest; in particular, the compactification in which the internal space has the topology of a seven–sphere, S^7, has been widely studied. A major problem, however, is that the theory is unable to account for the observed chirality of fermions in four dimensions.

It is necessary to construct a theory in which the Lorentz group has chiral representations. This is only possible if the theory has an even number of dimensions. For instance, chirality is a feature of six–dimensional Maxwell–Einstein theory compactified into $M^4 \times S^2$, where S^2 is a two–sphere. The Maxwell field is assumed to have a magnetic monopole configuration on the two–sphere, and it is this that acts as a source of chirality in the reduced theory [37]. Salam and Sezgin [11] have extended this theory to six–dimensional N=2 Maxwell–Einstein supergravity and found that both the spontaneous compactification into $M^4 \times S^2$ and the chirality coupling are retained here. Both theories involve a cosmological term and a real scalar field in six dimensions. The supergravity theory, however, is the more natural in that the cosmological term is not introduced by hand. Rather, it is a requirement of local supersymmetry and this determines both the sign and magnitude of the term. In fact, the term is not constant as previously, but is a function of the scalar field. Further details of the Salam-Sezgin model will be given in the following sections of this chapter where it is employed to study the effects of decoherence. It should be stressed here that the theory is used only as a toy model; it is a simple model which is relatively easy to work with but which fails to be realistic. It does, however, have the virtue of displaying many of the chacteristics of ten–dimensional supergravity, especially chirality and a vanishing cosmological constant in four dimensions. Since ten–dimensional supergravity can be derived from superstrings it is of particular interest.

It is possible to regard the extra spatial dimensions of Kaluza–Klein theories merely as a mathematical tool employed to bring about the desired unification. A more fruitful and satisfactory approach is to regard these dimensions as a part of physical reality. One can then look to cosmology as a test of the theories. In particular, the behaviour of the very early universe in higher dimensions is of interest since it is expected that in the Planckian era all dimensions would have been of the same scale. Then the extra dimensions would play a role going beyond their more indirect effect on particle interactions at lower energies.

One effect is the production of massive particles at high temperatures which are then unable to decay, leading to a cosmological abundance problem. For instance, massless scalar fields in the five–dimensional theory can give rise to massive modes in four dimensions. The Klein–Gordon equation in five dimensions is

$$\left(\Box + \frac{1}{R^2}\frac{\partial^2}{\partial\theta^2}\right) \phi\,(x,\theta) = 0\ , \tag{1.3.9}$$

where $\Box$ is the d'Alembertian operator in four dimensions. Since the field can be Fourier expanded as in (1.3.5), the equations for the components are given by

$$\left(\Box - m_n^2\right)\, \phi^n(x) = 0 \tag{1.3.10}$$

where

$$m_n = \frac{n}{R}. \tag{1.3.11}$$

Thus, since the radius of the internal space is small, (1.3.8), all the fields $\phi^n(x)$ will have masses of the order of the Planck mass. The only exception is the n=0 mode which is massless. The behaviour is similar in more general theories. At energies significantly lower than the compactification scale, the non–zero modes do not contribute to the dynamics, and the zero modes alone produce the low mass particle spectrum. However, at temperatures comparable to the compactification scales, the massive modes will be excited, and it is important to track their decay in the evolving universe. Since the Fourier expansion (1.3.5) involves an infinite number of components, there are infinite 'towers' of these modes, earning them the name 'pyrgons' [38]. In many models the pyrgons are stable and can only be removed from the universe by annihilation with antipyrgons. However, the expansion of the universe makes annihilation ineffective for those modes of mass less than the Planck mass. It is reasonable to assume that in the pre–compactification universe, the number density of pyrgons is equal to that of photons. This severly restricts the Kaluza–Klein models which are realistically available, although it has been suggested [38] that an entropy producing process, such as inflation, would remove the problem.

Another constraint on the suitability of models comes from the consideration of the time dependence of the fundamental constants. In Kaluza–Klein theories it is the higher dimensional constants which are fundamental; the four–dimensional effective constants vary with the expansion or contraction of the internal space. Thus the observation that, for instance, the gravitational constant has negligible time dependence implies that the radius of the internal space must now be constant. This is reasonable since the construction of Kaluza–Klein theories requires the higher dimensions to be small. Then the only alternative to small oscillations about a constant radius is contraction, which would lead to singularities. To be successful, therefore, a model must allow for a constant internal space at late times. This is achieved in the Salam–Sezgin model by the monopole configuration on the two–sphere.

Classically, cosmological studies have centred on the time evolution of the radii of the two spaces in the theory in an attempt to understand the disparity in their sizes. Typically, the approach is to assume that the (4+d)–dimensional metric is the sum of the usual Robertson–Walker metric in four dimensions, with scale factor $a(t)$, and the metric of a d–sphere, with time dependent scale factor $b(t)$. The general form of the energy–momentumtensor for an isotrpic universe is then

$$T_{00} = \rho; \qquad T_{ij} = p g_{ij}; \qquad T_{mn} = p' g_{mn} \,. \tag{1.3.12}$$

The Einstein equations provide equations for the scale factors. One then searches for cosmological solutions fulfilling the demand that the internal space is stable and of constant radius whilst the rest of the universe is expanding as observed.

Cosmologies of the Salam–Sezgin model have been studied by Maeda and Nishino [39], Halliwell [40] and Lonsdale [41]. The model involves many fields, but at early times the fermionic fields can be neglected. The simplified action is then of the form

$$S = \int \left[-\frac{{}^6R}{4\kappa^2} - \frac{1}{2}\partial_\alpha \Phi \partial^\alpha \Phi - \frac{1}{4} e^{\sqrt{2}\kappa\Phi} F_{\alpha\beta}F^{\alpha\beta} - \frac{g^2}{2\kappa^4} e^{-\sqrt{2}\kappa\Phi} \right] \sqrt{-{}^6g}\, d^6x \ , \quad (1.3.13)$$

where $\kappa^2/4\pi$ is the gravitational constant in six dimensions, ${}^6R$ is the curvature of the full six–dimensional space, g is a coupling constant, $\Phi(x)$ is a scalar field, and $F_{\alpha\beta}$ is a six–dimensional rank two field strength. The homogeneity and isotropy of the observed universe restricts the field strength to be non–zero on the internal space only. Here the field is taken to be a monopole of integral charge m. Thus

$$F_{ab} = \frac{m}{2g} dx^a \wedge dx^b \quad (1.3.14)$$

with all other components zero.

Salam and Sezgin [11] have shown that the field equations in flat space fix the monopole charge as $m = \pm 1$. Under this configuration the compactification into the product of a four–dimensional spacetime with a two–sphere is realised. The six–dimensional metric can then be expressed as

$$ds^2 = -N^2(t)\, dt^2 + a^2(t)\, d\Omega_3^2 + b^2(t)\, d\omega_2^2 \ , \quad (1.3.15)$$

where $N(t)$ is the time dependent lapse function and $a(t)$ and $b(t)$ are time dependent scale factors for the three– and two–dimensional spaces respectively. $d\omega_2^2$ is the metric on a unit two–sphere, whilst $d\Omega_3^2$ is the spatial part of a Robertson–Walker metric. i.e.

$$d\Omega_3^2 = \frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) . \quad (1.3.16)$$

Maeda and Nisino [39], by examining the equations of motion for the parameters N, a and b, showed that the model admits two cosmological solutions. Concentrating on the case where $k = 0$, they found two possible families of solution. One family coresponds to the universe being in a vacuum state and includes the solution (Minkowski space)$\times$(constant S^2). However, this solution is shown to be unstable to perturbations in the initial conditions. The other possible family of solutions describes a universe dominated by four–dimensional radiation, and all solutions of this family approach (Friedmann universe)×(constant S^2) regardless of initial conditions.

The case of $k = 1$ has been studied by Lonsdale [41]. He solved the equations of motion for the internal radius and the scalar field numerically and found that their late time behaviour was independent of initial conditions; both exhibited damped oscillations about some constant value. At the same time the scale factor of the observable universe undergoes straightforward expansion with

$$a(t) \sim t \, . \tag{1.3.17}$$

Thus the solution (Friedmann)×(constant S^2) is naturally realised.

Halliwell [40] also considered the behaviour of the internal radius and the scalar field, but did so using phase diagrams. He found that the $k = -1$ trajectorieswere attracted towards the $k = 0$ trajectory. Thus an inflationary mechanism exists through which spatial flatness is attained. However, the mechanism fails to solve the other problems which inflation usually attempts to resolve, such as the horizon problem. The $k = +1$ trajectories move away from the $k = 0$ trajectory during the expansion era and therefore do not seem to be of physical interest.

1.4 Quantum Kaluza–Klein Theories

In all these classical studies of Kaluza–Klein theories, there always remains the problem of our ignorance about the initial conditions of the universe. Quantum cosmology suggests a resolution of this problem through the adoption of the Hartle–Hawking proposal. The extension of this proposal to higher dimensions should therefore give a more complete description of Kaluza–Klein cosmologies. By analogy with the four–dimensional situation, in (4+d) dimensions the wavefunction Ψ is a functional of the (4+d-1)–metric h_{IJ} and matter fields Φ on a (4+d-1) surface S. It is defined by the path integral over all matter fields and compact (4+d)–metrics which match h_{IJ} on S [42]. Unlike in four dimensions, any given (4+d-1) space may not form the boundary of a (4+d) space. To ensure that this is the case, and that the wavefunction is non–zero, it may be necessary to consider S to be disconnected.

Using this generalization of the Hartle–Hawking prescription, both Lonsdale and Halliwell have calculated the wavefunction for the Salam–Sezgin model in the minisuperspace approximation. Since only compact metrics can be considered it is necessary to take $k = 1$.

Lonsdale [41] took $\Psi = 1$ as the boundary condition in the Wheeler–De Witt equation and then solved numerically to find the wavefunction at larger a. Three distinct regions of behaviour were found: one lying along the line defining the minimum of the potential of the four-dimensional effective action, and the other two lying either side of this line. The wavefunction exhibits oscillations about this line; on one side the oscillations are exponentially damped, whilst on the other they are not – the different behaviour being due to the asymmetry of the potential. These oscillations represent universes with Lorentzian metric, and in the WKB approximation imply that classical trajectories originate away from the minimum of the potential and evolve towards it.

Halliwell [40], working analytically, found generally similar behaviour, although where Lonsdale had found exponentially damped oscillations, Halliwell found pure exponential behaviour. Thus, in this region the wavefunction corresponds to a Euclidean six–geometry and is therefore classically forbidden.

In this way, it is seen that in the classical limit the wavefunction picks out a particular solution thereby providing initial conditions for the universe.

1.5 Decoherence in the Salam–Sezgin Model

The action and Hamiltonian

The line of minima in the potential of the action described above means that the scale of the compactified internal space, although fixed, is arbitrary. This is a feature common to many supersymmetric models. The universe can exist in a number of different states corresponding to the different possible scales at which the extra dimensions become compactified. It is therefore necessary to ensure that quantum interference between these different states is suppressed. Thus the notion of decoherence is introduced. The aim here is to illustrate that in the Salam–Sezgin model the differently scaled internal spaces do indeed decohere. In order to do this using the approach advocated by Zurek and described in §1.2 above, it is necessary to identify an environment. Since the previous quantum cosmological studies of the Salam–Sezgin model by Lonsdale and Halliwell have concentrated on the homogeneous modes of the minisuperspace approximation, an obvious candidate for the environment is the inhomogeneous modes of the internal space and the scalar field. An extension of minisuperspace to include inhomogeneous modes in four dimensions has been carried out by Halliwell and Hawking [43].

As above, the Salam–Sezgin action is taken as

$$S = \int \left[-\frac{{}^6R}{4\kappa^2} - \frac{1}{2}\partial_\alpha \Phi \partial^\alpha \Phi - \frac{1}{4} e^{\sqrt{2}\kappa\Phi} F_{\alpha\beta} F^{\alpha\beta} - \frac{g^2}{2\kappa^4} e^{-\sqrt{2}\kappa\Phi} \right] \sqrt{-{}^6g}\, d^6x \ , \qquad (1.5.1)$$

with the field strength given by

$$F_{ab} = \frac{m}{2g} dx^a \wedge dx^b; \qquad \text{all other components zero} . \qquad (1.5.2)$$

Again, the presence of this monopole on the internal space induces the compactification with topology $M^4 \times S^2$. Here, M^4 is taken to be a Friedmann–Robertson–Walker universe with $k = 1$ and scale factor $a(t)$, and S^2 is a perturbed two–sphere; the perturbations providing the environment modes.

Thus the metric is given by

$$ds^2 = e^{-2u(t)} \left(-N^2(t)\, dt^2 + a^2(t)\, d\omega_3^2 + b^2 e^{4u(t)} \left(d\omega_2^2 + h_{ab} dx^a dx^b \right) \right) \qquad (1.5.3)$$

where $d\omega_3^2$ is the metric on a unit three–sphere and $d\omega_2^2$ is the metric on a unit two–sphere. The time variation of the radius of the internal space has been entered explicitly and is given by $u(t)$, whilst b is a constant length parameter.

The metric has been Weyl rescaled by a factor $e^{-2u(t)}$ to remove some cross terms from the action. This also ensures that the coefficient of the four–dimensional Ricci scalar in the action has no time dependence; that is the four–dimensional gravitational constant, G, is truely a constant. The gravitational constants in six and four dimensions are related by

$$\kappa^2/4\pi = 4\pi b^2 G \ . \qquad (1.5.4)$$

The perturbation of the internal space is given by h_{ab}, which may be expanded in terms of harmonics on the two–sphere:

$$h_{ab} = \sum_n \left(\frac{1}{2} \bar{g}_{ab} \alpha_n Q_n + \beta_n P_{n\,ab} + \gamma_n S_{n\,ab} \right) \tag{1.5.5}$$

with the angular momentum labels l and m suppressed. Here, $\bar{g}_{ab}$ is the metric of the unperturbed two–sphere, and $\alpha_n, \beta_n, \gamma_n$ are coefficients dependent on time t only. $Q_n(x^a)$ are the scalar harmonics on the two–sphere, $P_{n\,ab}(x^c)$are derived from the scalar harmonics by the equation

$$P_{n\,ab} = \frac{1}{l\,(l+1)} Q_{n|ab} + \frac{1}{2} \bar{g}_{ab} Q_n \ , \tag{1.5.6}$$

where a vertical bar indicates covariant differentiation with respect to the unperturbed two–sphere. $S_{n\,ab}(x^c)$ are the odd and even traceless tensor harmonics derived from the derivatives of the transverse vector harmonics $S_{n\,a}(x^b)$ such that

$$S_{n\,ab} = S_{n\,a|b} + S_{n\,b|a} \qquad \text{with} \qquad S_a^{\ |a} = 0 \ . \tag{1.5.7}$$

The transverse, traceless tensor harmonics are neglected since they have no physical content on the two–sphere. Further details of the complete set of harmonics on the two–sphere, their eigenvalue equations and the normalizations adopted here, are given in the appendix.

In a similar manner, the scalar field and lapse function are both expressed in terms of the scalar harmonics:

$$N = N_0 \left[1 + \sum_n \delta_n Q_n \right] , \tag{1.5.8}$$

$$\Phi = \phi\,(t) + \sum_n \epsilon_n Q_n \ , \tag{1.5.9}$$

where δ_n, ϵ_n are coefficients again dependent on time only.

In the following calculation, perturbative terms are retained to second order only.

Since the metric has an overall conformal rescaling, the action (1.5.1) can be rewritten

$$S = \int \Big[-\frac{1}{4\kappa^2}\left(e^{2u}\,{}^6\tilde{R} + 5e^{2u}\tilde{\nabla}^2 u\right) - \frac{1}{2}\partial_\alpha \Phi \partial^\alpha \Phi \\ - \frac{m^2}{8g^2} b^{-4} e^{-4u} e^{\sqrt{2}\kappa\Phi} - \frac{g^2}{2\kappa^4} e^{-\sqrt{2}\kappa\Phi} \Big] e^{-6u} \sqrt{-{}^6\tilde{g}}\, d^6x \, , \tag{1.5.10}$$

where $\tilde{\nabla}$ and $\sqrt{-{}^6\tilde{g}}$ are the covariant derivative and the measure respectively, evaluated over the full six–dimensional, perturbed, but not rescaled, space. The curvature term here is also evaluated over the perturbed, unscaled space. It can be expanded in terms of the curvature, ${}^6\bar{R}$, on the unperturbed six–dimensional space with metric

$$ds^2 = -dt^2 + a^2(t)\, d\omega_3^2 + b^2 e^{4u(t)}\, d\omega_2^2 \, . \tag{1.5.11}$$

Then the curvature is given by

$$\sqrt{-{}^6\tilde{g}}\,{}^6\tilde{R} = \sqrt{-{}^6\bar{g}} \Big\{ {}^6\bar{R} - \left({}^6\bar{R}^{\alpha\beta} - \frac{1}{2}\bar{g}^{\alpha\beta}\,{}^6\bar{R}\right)\hat{h}_{\alpha\beta} \\ - \frac{1}{4}\hat{h}^{\alpha\beta}\left(-\bar{g}_{\alpha\gamma}\bar{g}_{\beta\delta}\bar{\nabla}^2 - 2\,{}^6\bar{R}_{\alpha\gamma}\bar{g}_{\beta\delta} - 2\,{}^6\bar{R}_{\alpha\beta\gamma\delta} + {}^6\bar{R}\bar{g}_{\alpha\gamma}\bar{g}_{\beta\delta}\right)\hat{h}^{\gamma\delta} \\ + \frac{1}{8}\bar{g}^{\alpha\beta}\hat{h}^\epsilon_\epsilon\left(-\bar{g}_{\alpha\gamma}\bar{g}_{\beta\delta}\bar{\nabla}^2 - 2\,{}^6\bar{R}_{\alpha\gamma}\bar{g}_{\beta\delta} - 2\,{}^6\bar{R}_{\alpha\beta\gamma\delta} + {}^6\bar{R}\bar{g}_{\alpha\gamma}\bar{g}_{\beta\delta}\right)\hat{h}^{\gamma\delta} \\ + \frac{1}{2}\bar{\nabla}^\gamma\left(\hat{h}_{\gamma\alpha} - \frac{1}{2}\bar{g}_{\gamma\alpha}\hat{h}^\beta_\beta\right)\bar{\nabla}^\delta\left(\hat{h}^\alpha_\delta - \frac{1}{2}\bar{g}^\alpha_\delta\hat{h}^\epsilon_\epsilon\right)\Big\}. \tag{1.5.12}$$

Horizontal bars over the symbols indicate that these quantities are evaluated using the metric (1.5.11), and

$$\hat{h}_{ab} = b^2 e^{4u} h_{ab}; \qquad \hat{h}^{ab} = b^{-2} e^{-4u} h^{ab}; \qquad \hat{h}^a_b = h^a_b \tag{1.5.13}$$

with all other components zero. Also, the measure can be expanded

$$\sqrt{-{}^6\tilde{g}} = \sqrt{-{}^4g}\sqrt{-{}^2\bar{g}}\left(1 + \frac{1}{2}h^a_a + \frac{1}{8}\left(h^a_a\right)^2 - \frac{1}{4}h^a_b h^b_a + \ldots\right) \tag{1.5.14}$$

and

$$\tilde{\nabla}^2 u = -a^{-3} e^{-4u}\left(1 + \frac{1}{2}h^a_a + \frac{1}{8}(h^a_a)^2 - \frac{1}{4}h^a_b h^b_a\right)^{-1} \cdot \\ \cdot \frac{\partial}{\partial t}\left[a^3 e^{4u}\left(1 + \frac{1}{2}h^a_a + \frac{1}{8}(h^a_a)^2 - \frac{1}{4}h^a_b h^b_a\right)\frac{\partial u}{\partial t}\right] . \tag{1.5.15}$$

The above expressions, together with the harmonic expansions (1.5.5), (1.5.8), (1.5.9) enable one to write down the Lagrangian, defined by

$$S = \frac{1}{16\pi G}\int L\,dt\,. \tag{1.5.16}$$

The integration over the three spatial dimensions of the Friedmann universe is straight forward. The integration over the internal space is given by the properties of the harmonics on the two–sphere, as described in the appendix.

The Lagrangian is found to be given by

$$\begin{aligned} L = \frac{\pi a^3}{2N_0}\Big\{ &\left[-6\left(8\pi+2\delta_n^2\right)+\tfrac{3}{2}\left(\alpha_n^2+C^s\beta_n^2+C^v\gamma_n^2\right)+6\delta_n\alpha_n\right]a^{-2}\dot{a}^2 \\ &+\left[4\left(8\pi+2\delta_n^2\right)+3\left(\alpha_n^2+C^s\beta_n^2+C^v\gamma_n^2\right)+28\delta_n\alpha_n-7\alpha_n^2\right]\dot{u}^2 \\ &-\tfrac{1}{4}\dot{\alpha}_n^2+\tfrac{1}{4}C^s\dot{\beta}_n^2+\tfrac{1}{4}C^v\dot{\gamma}_n^2+3C^s\beta_n a^{-1}\dot{a}\dot{\beta}_n+3C^v\gamma_n a^{-1}\dot{a}\dot{\gamma}_n \\ &+6\delta_n a^{-1}\dot{a}\dot{\alpha}_n+\left[4\delta_n+4\alpha_n\right]\dot{u}\dot{\alpha}_n+2C^s\beta_n\dot{u}\dot{\beta}_n+2C^v\gamma_n\dot{u}\dot{\gamma}_n \\ &+2\kappa^2\left[8\pi+\tfrac{1}{4}\alpha_n^2-\tfrac{1}{4}\left(\alpha_n^2+C^s\beta_n^2+C^v\gamma_n^2\right)+2\delta_n^2\right]\dot{\phi}^2 \\ &+4\kappa^2\dot{\epsilon}_n^2+8\kappa^2\left[\frac{\alpha_n}{2}-\delta_n\right]\dot{\epsilon}_n\dot{\phi}\Big\}+V'\,, \end{aligned} \tag{1.5.17}$$

where V' is the total potential of the perturbed model:

$$\begin{aligned} V' = \tfrac{1}{2}\pi N_0 a^3\Big\{ &3a^{-2}\left[16\pi+2\delta_n\alpha_n-\tfrac{1}{2}C^s\beta_n^2-\tfrac{1}{2}C^v\gamma_n^2\right] \\ &+b^{-2}e^{-4u}\left[16\pi-2\delta_n\alpha_n-\tfrac{1}{2}\left(l^2+l+2\right)\alpha_n^2-\tfrac{1}{4}\left(l^2+l-3\right)C^v\gamma_n^2\right] \\ &-2\kappa^2 b^{-2}e^{-4u}l(l+1)\epsilon_n^2 \\ &+4\kappa^2 e^{-2u}\Big[-\frac{n^2}{8\kappa g^2}b^{-4}e^{-4u}e^{\sqrt{2}\kappa\phi}\big(8\pi+\alpha_n\delta_n+2\sqrt{2}\kappa\delta_n\epsilon_n+\tfrac{1}{4}\alpha_n^2 \\ &\qquad-\tfrac{1}{4}\left(\alpha_n^2+C^s\beta_n^2+C^v\gamma_n^2\right)+\sqrt{2}\kappa\alpha_n\epsilon_n+2\kappa^2\epsilon_n^2\big) \\ &\qquad-\frac{g^2}{2\kappa^3}e^{-\sqrt{2}\kappa\phi}\big(8\pi+\alpha_n\delta_n-2\sqrt{2}\kappa\delta_n\epsilon_n+\tfrac{1}{4}\alpha_n^2 \\ &\qquad-\tfrac{1}{4}\left(\alpha_n^2+C^s\beta_n^2+C^v\gamma_n^2\right)-\sqrt{2}\kappa\alpha_n\epsilon_n+2\kappa^2\epsilon_n^2\big)\Big]\Big\}\,. \end{aligned} \tag{1.5.18}$$

The Hamiltonian is of the form $H = p_i\dot{q}^i - L$ where p_i are the canonical momenta conjugate to co–ordinates q^i, such that $p_i = \partial L/\partial\dot{q}^i$. This gives

$$H = N_0\left\{H_0+\frac{1}{\pi a^3}\sum_n \delta_n H_n^{(1)}+\frac{a^3}{2\kappa^2\pi}\sum_n H_n^{(2)}\right\} \tag{1.5.19}$$

where H_0 is the Hamiltonian of the corresponding unperturbed model:

$$H_0 = \frac{1}{\pi a^3}\left\{-\frac{1}{12}\frac{a^2 p_a^2}{8\pi} + \frac{1}{8}\frac{p_u^2}{8\pi} + \frac{1}{4\kappa^2}\frac{p_\phi^2}{8\pi} + V(a,u,\phi) - 24\pi^3 a^4\right\} . \qquad (1.5.20)$$

$H_n^{(1)}$ is the first order perturbation to the Hamiltonian and $H_n^{(2)}$ is the second order perturbation.

The potential term $V(a,u,\phi)$ takes the form

$$V(a,u,\phi) = 4\pi^3 a^6 b^{-2} e^{-4u}\left[\frac{m^2}{2g}\kappa^2 b^{-2} e^{-2u} e^{\sqrt{2}\kappa\phi} + \frac{2g^2}{\kappa^2} b^2 e^{2u} e^{-\sqrt{2}\kappa\phi} - 2\right] . \qquad (1.5.21)$$

At large u, the potential tends asymptotically to zero. This represents a situation where the internal space has an arbitrarily large radius and so the topology of the metric approaches M^6. The position of the minimum of the potential, which represents a space with topology $M^4 \times S^2$, fixes the relationship between the scale u and the field ϕ, but beyond that the scale of the compactified internal space is arbitrary.

It should be noted that if the monopole on the internal space has unit charge, i.e. $m = \pm 1$, then the potential takes the form

$$V(a,u,\phi) = 16\pi^3 a^6 b^{-2} e^{-4u} \sinh^2(u - \phi') , \qquad (1.5.22)$$

where

$$\phi' = \frac{\kappa\phi}{\sqrt{2}} - \frac{1}{2}\ln\frac{2b^2 g^2}{\kappa^2} . \qquad (1.5.23)$$

This form of the potential is plotted as a function of u in figure 1.1. It can be seen that the potential is zero at its minimum. This corresponds to zero cosmological constant and is therefore of particular interest. Thus $|m|$ is set equal to one for the remainder of the calculation. The minimum now lies along a line defined by $u = \phi'$, so that close to the minimum ϕ may be regarded approximately as a function of u. Since the classical solutions exhibit heavily damped oscillations about the line of minima [41], the field ϕ need not be considered independently in the calculation of the density matrix.

By suitable choice of the lapse formation δ_n, the zero energy Schrödinger equation may be separated into homogeneous and inhomogeneous parts. Thus one obtains the Wheeler – De Witt equation

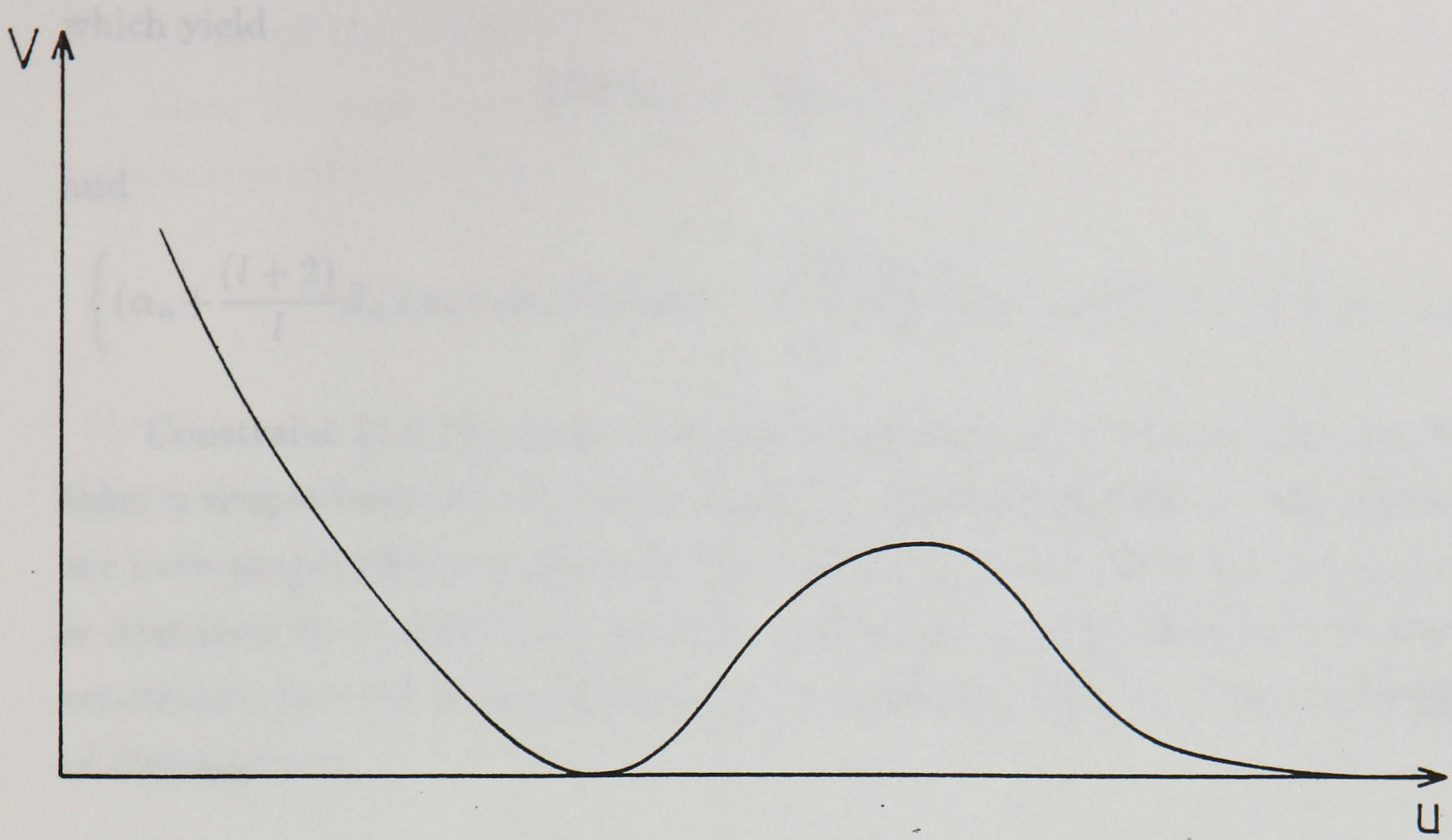

Fig 1.1 A sketch of the potential V.

$$\left\{ H_0 + \frac{a^3}{2\kappa^2\pi} \sum_n H_n^{(2)} \right\} \Psi = 0 \qquad (1.5.24)$$

and the Hamiltonian constraint

$$H_n^{(1)} \Psi = 0 \,. \qquad (1.5.25)$$

In (1.5.25) we are free to make the gauge choice $\delta_n = 0$, in which case the constraint (1.5.25) reduces to

$$\begin{aligned} \Bigg\{ \frac{\alpha_n}{(8\pi)^2} \Big[\frac{a^2 p_a^2}{12} + \frac{15}{8} p_u^2 + \frac{p_\phi^2}{2} - a p_a p_u \Big] + \frac{1}{8\pi} \Big[2 a p_a p_{\alpha_n} + 2 p_u p_{\alpha_n} - \frac{1}{2\kappa^2} p_{\epsilon_n} p_\phi \Big] \\ + \alpha_n \big[3\pi^2 a^4 + \pi^2 a^6 b^{-2} e^{-4u} (\cosh(u - \phi') - 1) \big] \\ - \epsilon_n 2\sqrt{2} \kappa \pi^2 a^6 b^{-2} e^{-4u} \sinh(u - \phi') \Bigg\} \Psi = 0 \,. \end{aligned} \qquad (1.5.26)$$

In addition, one has the momentum constraints [24]

$$\left\{ (\bar{g}^{-\frac{1}{2}} p^{\alpha\beta})_{;\beta} - \bar{g}^{\frac{1}{2}} \bar{g}^{\alpha\beta} \frac{\partial \Phi}{\partial x^\beta} p_\Phi \right\} \Psi = 0 \,, \qquad (1.5.27)$$

which yield

$$\{32\pi p_{\gamma_n} + \gamma_n(p_u + ap_a)\}\Psi = 0 \tag{1.5.28}$$

and

$$\left\{(\alpha_n+\frac{(l+2)}{l}\beta_n)(p_u+ap_a)+64\pi p_{\alpha_n}-32\pi\frac{(l+2)}{l}p_{\beta_n}+8b^2e^{2u}\epsilon_n p_\phi\right\}\Psi = 0\ . \tag{1.5.29}$$

Constraint (1.5.28) shows that the dependence of the wave function upon γ_n takes a simple form and we may evaluate the wave function at $\gamma_n = 0$. These modes are pure gauge degrees of freedom. The behaviour of the scalar modes α_n, β_n and ϵ_n is described by (1.5.26) and (1.5.27). In principle it is possible to substitute these constraints into the homogeneous part of the Hamiltonian to obtain a Hamiltonian of diagonal form

$$H = N_0\left\{H_0 + \frac{a^3}{2\kappa^2\pi}\sum_n\left(\frac{1}{2}a^{-6}p^2_{\zeta_n} + \frac{1}{2}\omega_n^2\zeta_n^2\right)\right\}, \tag{1.5.30}$$

where ζ_n is a suitable combination of $\alpha_n, \beta_n, \epsilon_n$ and their conjugate momenta. In general however, this approach leads to a complicated expression for the frequency ω_n. It is simpler to make the gauge choice $\alpha_n = \beta_n = 0$ initially, and then to find the Hamiltonian. Under this gauge choice, the constraints are consistently solvable if the calculation is restricted to the limit of p_u and p_ϕ both tending to zero. This is permissable since the era of most interest to us as observers today is of course at late times when the compactification radius of the internal space is fixed. To study the decoherence between different, but fixed, compactifications means to concentrate on the minimum of the potential well where the momenta conjugate to both u and ϕ are zero.

In this limit, the second order perturbation of the Hamiltonian represents a series of simple harmonic oscillators of non-constant mass,

$$H_n^{(2)} = \frac{1}{2}a^{-6}p^2_{\epsilon_n} + \frac{1}{2}\omega_n^2\epsilon_n^2 \tag{1.5.31}$$

where the frequencies are given by

$$\omega_n^2 = 4\kappa^4\pi^2b^{-2}e^{-4u}(l^2 + l + 2)\ . \tag{1.5.32}$$

Solutions of the Wheeler–De Witt equation

Since the perturbation modes are not coupled to each other, it is possible to assume separable solutions

$$\Psi = \Psi_0(a,u)\prod_n \Psi_n(a,u,\epsilon_n) \tag{1.5.33}$$

where Ψ_0 is assumed classical and described by the WKB approximation such that $\Psi_0 = Ae^{iS}$, where A is a slowly varying function of the variables. Ψ_0 behaves as described by Lonsdale [41], exhibiting a wavelike motion travelling towards the line $u = \phi'$ where it is highly peaked. The Wheeler–De Witt (1.5.24) gives

$$H_0(\Psi_0).\prod_n \Psi_n + \frac{1}{8\pi^2 a^3}\Psi_0\left(-\frac{a^2p_a^2}{12} + \frac{p_u^2}{8} + \frac{p_\phi^2}{4\kappa^2}\right)\prod_n \Psi_n + \frac{a^3}{2\kappa^2\pi}\Psi_0 \sum_n H_n^{(2)}\left(\prod_n \Psi_n\right)$$
$$+ \frac{2}{8\pi^2 a^3}\left[-\frac{a^2 p_a}{12}(\Psi_0)p_a\left(\prod_n \Psi_n\right) + \frac{p_u}{8}(\Psi_0)p_u\left(\prod_n \Psi_n\right) + \frac{p_\phi}{4\kappa^2}(\Psi_0)p_\phi\left(\prod_n \Psi_n\right)\right] = 0\ . \tag{1.5.34}$$

The second term in (1.5.34) may be neglected by assuming the adiabatic approximation in which $a(t)$ is slowly varying. This implies that the dependence of Ψ_n on $a(t)$ and $u(t)$ is negligible compared with the dependence of Ψ_0 on these variables.

Furthermore, the semiclassical approximation enables the third term to be simplified, since for classical Ψ_0, one has, for example

$$p_a\Psi_0 = \Psi_0\frac{\partial S}{\partial a} = -\Psi_0 48\pi^2 a\dot{a}\ . \tag{1.5.35}$$

Thus the last term in (1.5.34) becomes

$$\Psi_0[\dot{a}p_a + \dot{u}p_u + \dot{\phi}p_\phi]\prod_n \Psi_n = -\Psi_0 i\frac{d}{dt}\left(\prod_n \Psi_n\right) \tag{1.5.36}$$

regarding the momenta as quantum operators on Ψ_n, and where t is the time coordinate as defined by the metric (1.5.3).

Then (1.5.34) can be written as

$$\left(H_0 + \sum_n E_n\right)\Psi_0 = 0 \tag{1.5.37}$$

where the E_n are given by

$$-i\frac{d\Psi_n}{dt} + \frac{a^3}{2\kappa^2\pi}H_n^{(2)}\Psi_n = E_n\Psi_n \ . \tag{1.5.38}$$

The vacuum energy $\sum_n E_n$ should be taken into account as a back reaction term added to the original potential. In the present context of a super-symmetric theory, this term will have cancellations between the fermionic and bosonic contributions. If, in spite of such cancellations, this term does not vanish at the minimum of the potential, then some additional mechanism has to be invoked to explain the vanishing cosmological constant.

Since the perturbation dependent parts of the Hamiltonian, H_n, have a simple harmonic form, they may be expressed in terms of annihilation operators A_n such that

$$H_n = A_n^\dagger A_n + \frac{1}{2} \tag{1.5.39}$$

which requires that

$$A_n = \frac{1}{\sqrt{2}}(\omega_n^{1/2}a^{3/2}\epsilon_n + ia^{-3/2}\omega_n^{-1/2}p_{\epsilon_n}) \ . \tag{1.5.40}$$

These are the annihilation operators for the eigenfunctions of Ψ_n. Thus Ψ_n may be expanded as a series

$$\Psi_n = \sum_n C_{nN}f_n, \tag{1.5.41}$$

of oscillator wavefunctions f_n,

$$\frac{a^3}{2\kappa^2\pi}H_n^{(2)}f_N = E_{nN}f_N \ . \tag{1.5.42}$$

The ground state function, f_0, is given by

$$f_0 = (\frac{\omega_n a^3}{\pi})^{\frac{1}{4}}e^{-\frac{1}{2}\omega_n a^3\epsilon_n^2} \ . \tag{1.5.43}$$

The coefficients, C_{nN}, obey the equation

$$\frac{dC_{nN}}{dt} = i\,(E_n - E_{nN}) - \sum_M C_{nM}\int\frac{df_M}{dt}f_N\,d\epsilon_n \ . \tag{1.5.44}$$

The interaction term, $\int \frac{df_M}{dt} f_N \, d\epsilon_n$, represents transitions between states cused by expansion of the scale factor. At the late times of interest here, this can be neglected. However, retention of this term in an analysis at early times, would enable one to study pyrgon production in the early universe.

If $(E_n - E_{nN})$ is constant, the perturbation dependent modes of the wave function are given by

$$\Psi_n = e^{iE_n t} \sum_N C_{nN}^{(0)} e^{-iE_{nN} t} \frac{\left(A^\dagger\right)^N}{\sqrt{N!}} f_0 \,, \tag{1.5.45}$$

where $C_{nN}^{(0)}$ is the value of the coefficient C_{nN} at $t = 0$. It will be assumed that the fields ϵ_n are initially in the ground state so that $C_{nN}^{(0)} = \delta_{N0}$. Thus N is set to zero in the remainder of the calculation.

The assumption that $C_{nN}^{(0)} = \delta_{N0}$ is equivalent to the assumption that the wavefunction is initially in the Hartle–Hawking state. The wavefunction (1.5.45) can be written in the general form

$$\Psi_L = \sum_N C_{nN}^{(0)} f_N \, e^{-iE_{nN} t} \,, \tag{1.5.46}$$

where $E_{nN} = (n + \frac{1}{2})\omega_n$. This describes its behaviour in the Lorentzian era. However, there was in the past a Euclidean era when the scale factor was less than some fixed value a_0. At this time the wavefunction can be described by

$$\Psi_E = \sum_N \hat{C}_{nN} f_N \, e^{-E_{nN}\tau} \,, \tag{1.5.47}$$

where τ is the Euclidean time. At $a = a_0$, these two wavefunctions must be matched as in refereneces [16,17]. If $\omega_n \tau_0 \gg 1$, as in the adiabatic approximation, then

$$\frac{|C_{nN}^{(0)}|}{|C_{n0}^{(0)}|} = \frac{|\hat{C}_{nN}|}{|\hat{C}_{n0}|} e^{-n\omega_n \tau_0} \ll 0 \tag{1.5.48}$$

providing $\frac{|\hat{C}_{nN}|}{|\hat{C}_{n0}|}$ is not big. This can be verified by imposing the Hartle–Hawking boundary condition. Then all coefficients $C_{nN}^{(0)}$ can be neglected except $C_{n0}^{(0)}$.

The density matrix

Having obtained an expression for the wavefunction of a Kaluza–Klein universe, we are now in a position to write down the corresponding density matrix. As explained in §1.2 above, the decoherence of quantum systems is seen in vanishing off-diagonal terms of the reduced density matrix. That is, summation over the environment modes, separate from the modes being observed, causes interference terms between different states to become negligible. In this case, the different states are internal spaces with different scale factor u, and the environment consists of the perturbation modes ϵ_n. Thus in the present example the reduced density matrix is given by

$$\rho(a_1,u_1;a_2,u_2) = \sum_{\epsilon_n} \Psi^*\Psi \tag{1.5.49}$$

$$= \Psi_0^*\Psi_0 \prod_n \sqrt{2}\left(\frac{\omega_{1n}^{1/2}a_1^{3/2}}{\omega_{2n}^{1/2}a_2^{3/2}} + \frac{\omega_{2n}^{1/2}a_2^{3/2}}{\omega_{1n}^{1/2}a_1^{3/2}}\right) . \tag{1.5.50}$$

The frequencies ω_n are given in (1.5.32). Writing $c = \omega^{1/2}a^{3/2}$, the product in the density matrix may be expressed as the exponential of a sum:

$$\rho(a_1,u_1;a_2,u_2) = \Psi_0^*\Psi_0 \exp(-\frac{1}{2}\sum_n \log\frac{1}{2}(\frac{c_1}{c_2}+\frac{c_2}{c_1})) , \tag{1.5.51}$$

where

$$\log\frac{1}{2}(\frac{c_1}{c_2}+\frac{c_2}{c_1}) = \log\frac{1}{2}\left(\frac{a_1^{3/2}}{a_2^{3/2}}e^{-(u_1-u_2)} + \frac{a_2^{3/2}}{a_1^{3/2}}e^{-(u_2-u_1)}\right) . \tag{1.5.52}$$

Since (1.5.52) is independent of l, and therefore of n, the sum is trivial. Thus,

$$\rho(a_1,u_1;a_2,u_2) = \Psi_0^*\Psi_0 \exp(-\frac{N}{2}\log\frac{(c_1^2+c_2^2)}{2c_1c_2}) . \tag{1.5.53}$$

Expanding the logarithm gives

$$\exp(-\frac{N}{2}\log\frac{(c_1^2+c_2^2)}{2c_1c_2}) = \exp\left(-\frac{N}{4}\frac{(c_1-c_2)^2}{c_1c_2}\right)\ \exp\left(-\frac{N}{4}\frac{(c_1-c_2)^4}{4c_1c_2}\right).... \tag{1.5.54}$$

Taking the limit as N tends to inifinty, (1.5.54) yields a selection function, $S(c_1-c_2)$, which is non-zero only at $c_1 = c_2$ where it is equal to unity. Thus one can set

$$a_1^{3/2}e^{-u_1} = a_2^{3/2}e^{-u_2} , \tag{1.5.55}$$

and

$$\rho(a_1, u_1; a_2, u_2) \sim S(c_1 - c_2)\ . \tag{1.5.56}$$

Thus the reduced density matrix is diagonalized, implying that there is no interference between states with different values of $a^{3/2}e^{-u}$.

One can expect that if the calculation were extended in a suitable way, for instance by summing the density matrix over the inhomogeneous modes on the three–sphere, then one could extract a selection function dependent on the internal radius u only. This would be in agreement with the work done by Kiefer [28], in which he shows that there is no interference between the different values of the scale factor, a, of a Friedmann universe in a four–dimensional minisuperspace model.

It is therefore illustated, that in the toy model of six–dimensional N=2 Einstein–Maxwell supergravity, all interference between the quantum states of the internal space is suppressed. This is concordant with our everyday experience of a classical four–dimensional universe.

CHAPTER TWO

BLACK HOLES AND GRAVITATIONAL INSTANTONS

2.1 Instantons

Gravitational instantons play an important role in the Euclidean approach to quantum gravity. They are a natural extension of the idea of instantons which arises in Yang–Mills theories. These are solutions to the Yang–Mills field equations formulated in Euclidean spacetime and displaying particle–like properties.

Euclidean spacetime is obtained by the analytic continuation of the Minkowskian system through the replacement of real time t with imaginary time τ, such that

$$\tau = it \; . \tag{2.1.1}$$

The motivation behind such a step is unclear within the confines of classical theory, but it proves to be a powerful tool in quantum field theory. If Feynman's sum–over–histories is followed through to field theories, then the vacuum to vacuum transition amplitude in the presence of an external source, J, can be represented by a path integral

$$\langle O_{out} | O_{in} \rangle^J \propto \int \mathcal{D}\phi \, e^{iS(\phi,J)} \; , \tag{2.1.2}$$

where$|\, O\rangle$ denotes the vacuum state. However, (2.1.2) is difficult to define since the integrand is oscillatory. This problem is avoided if the contour of integration is rotated in the complex plane; that is, if we change to imaginary time. Then the transition amplitude in Euclidean space is given by

$$\langle O_{out} | O_{in} \rangle^J_E \propto \int \mathcal{D}\phi \, e^{-I(\phi,J)} \; , \tag{2.1.3}$$

where I is the Euclidean action. The integrand is now well–behaved and is particularly amenable to semiclassical manipulation.

In the semiclassical limit, the functional integral is dominated by the stationary points of the action. The first–order functional derivative of the action vanishes for solutions of the classical equations of motion. Clearly, only solutions of finite action contribute to the path integral. Thus the integral is dominated by finite action solutions of the equations of motion; that is, by instanton contributions which interpolate between the vacuum states.

Instanton techniques arise in a number of areas in theoretical physics. In particular, they have been used effectively in quantum chromodynamics to solve several outstanding problems. For instance, one such problem was the U(1) problem, where the quantum chromodynamical description of hadronic physics has problems with it's approximate invariance under chiral SU(2)×SU(2). In the limit of perfect SU(2)×SU(2) symmetry, an additional chiral U(1) symmetry and associated conserved current appear in QCD. Since this symmetry is not manifest it must be spontaneously broken. But this would mean that there must be an associated massless Goldstone boson. What happens to this Goldstone boson is the essence of the U(1) problem. The solution to the problem lies in the fact that the axial current associated with the SU(2) symmetry has an anomolous term. Through coupling the instanton to this anomoly, 't Hooft [44] was able to show that chiral U(1) could be spontaneously broken without producing a Goldstone boson.

Another problem upon which it is hoped that instantons may shed light is that of quark confinement within hadrons. It is required that a successful theory such as QCD should be able to predict a long range attractive force between quarks. However, perturbation theory predicts a force falling off as $1/r^2$, as is usual in electrodynamics. Thus an alternative approach is needed, and in fact this has been one of the prime motivations behind the study of instantons.

Much attention has also been paid to the implications that instantons carry for the vacuum structure of gauge theories. A physical interpretation of instantons is that of a quantum mechanical tunnelling phenomenon inducing transitions between topologically distinct vacua. Here the power of the Euclidean approach, and hence of instanton methods, can clearly be seen. In a theory with, for instance, a double well potential one expects to find tunnelling effects from one well to the other. However, the presence of the potential barrier means that no classical paths can describe such effects. Changing to imaginary time has the result of inverting the potential, so that what were formerly minima are now maxima. Thus classical, albeit Euclidean, paths do now exist and a semiclassical approximation may be implemented. The boundary conditions in the functional integral for the transition through the potential barrier imply that suitable instantons are bounce solutions. The solution starts at, say, $x = x_0$ as $\tau \to -\infty$, bounces off the classical turning point at the end of the barrier and returns to $x = x_0$ as $\tau \to \infty$.

It is therefore possible to use instantons to study the decay of a false vacuum. The procedure was first applied in a thermodynamical context to look at condensation in simple models of first–order phase transitions. The model may represent a supersaturated solution, a superheated fluid, or, as in the analysis of Langer [45], a ferromagnet below the Curie temperature. These are metastable states where thermodynamic fluctuations can cause 'droplets' to form in a different state – that is, a region of crystal is nucleated in the supersaturated solution; a bubble of gas appears in the superheated fluid; or a region of, say, spin down appears in a ferromagnet with a background of spin up. Usually, the droplets or bubbles which form will be small, so that the bulk energy gained is less than the surface enegy lost. In this case, the bubble will shrink and disappear. However, occasionally a bubble will form which is sufficiently large that it is energetically favourable for expansion to occur. Langer studied this process through the analytic continuation to negative magnetization of a function whose real part represents the free energy of the system. He suggested that the imaginary part of this function could be identified with the rate of decay of the metastable state under some suitably chosen stochastic process.

This picture has a direct analogue in field theory, where quantum fluctuations can cause a bubble of true vacuum to form in a universe which is otherwise in a false vacuum, i.e. the fields are at local, but not global, minimum. If sufficiently large, the bubble will grow until it eventually fills the whole universe. The decay rate per unit volume of the false ground state has been computed using instanton methods by Coleman [13] and Callan and Coleman [14], but this work is restricted to flat spacetime.

Of greater cosmological interest is the occurrance of tunnelling in a curved spacetime. This is of importance in the inflationary scenario of the early universe in which the universe undergoes exponential expansion and approaches a de Sitter space, in which it remains until a phase transition occurs. The probability of this phase transition can be computed by seeking classical solutions to the Euclidean Einstein equations. The inclusion of the curvature of spacetime in the analysis is found to cause the quantum transition to occur everywhere simultaneously, taking the whole universe to a true vacuum with the topolgy of a four–sphere [15].

This de Sitter instanton has been given a different interpretation by Vilenkin [2,3]. As explained in chapter one, the canonical formulation of quantum gravity in terms of the Wheeler–De Witt equation governing the evolution of the wavefunction of the universe, begs the question of the boundary conditions. Several boundary conditions have been proposed, among them the suggestion that the big bang is the result of some quantum tunnelling event. For instance, Atkatz and Pagels [46] proposed that the universe originates by tunnelling from some static geometry. However, this description is unappealing, since if the initial static geometry has a finite tunnelling amplitude, then it could not have existed for an infinite time before the big bang. Thus one must ask: where did the initial geometry come from? By contrast, Vilenkin uses the de Sitter insanton to describe a universe tunnelling into existence from 'nothing'.

Spacetime is described by the Friedmann–Robertson–Walker metric:

$$ds^2 = -dt^2 + a^2(t)\left[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\right] , \tag{2.1.4}$$

where the scale factor $a(t)$ evolves as

$$a(t) = H^{-1}\cosh(Ht) , \tag{2.1.5}$$

with

$$H = (8\pi G\rho/3)^{\frac{1}{2}} . \tag{2.1.6}$$

These equations describe a universe contracting before $t = 0$, reaching its minimum size, $a_0 = H^{-1}$, at $t = 0$ and expanding thereafter. The analogy with a particle is clear, with $a(t)$ replacing the particle co–ordinate. The situation is akin to a particle bouncing off a potential barrier at $a = a_0$. However, quantum mechanically the particle can tunnel through such a barrier. Thus analogy implies that the universe might have originated in a similar quantum tunnelling event, emerging with initial conditions $a = H^{-1}$, $\dot{a} = 0$.

The Euclidean version of (2.1.4) and (2.1.5) describe a four–sphere S^4. Thus the instanton describes tunnelling into de Sitter space. However, unlike the case of a particle penetrating a barrier, the above solution does not approach an initial state at $\tau \to \pm\infty$. Rather, de Sitter space is compact, defined only for $|\tau| \leq \pi/2H$. Thus the instanton is interpreted as representing tunnelling fom nothing, from a state which has no classical spacetime.

The idea of tunnelling provided by gravitational instantons offers a physical picture for recent proposals [16-18] on the inclusion of complex contours in the path integral for the wavefunction of the universe.

In the conjectural arena of quantum gravity there lie many problems. Not least among these is the need to specify a contour of integration in the Euclidean path integral. This is non–trivial since the gravitational action is linear in the curvature and can therefore go negative, unlike the Yang–Mills action, for instance, which is positive definite. In the case of gravity then, the path integral may fail to converge if taken over real Euclidean metrics. Gibbons, Hawking and Perry [47] identified the conformal factor of the metric as being that part responsible for the action becoming negative and arbitrarily large. To sidestep this problem, they proposed a procedure for the evaluation of the path integral which essentially involved separating the functional space of metrics into conformal equivalence classes, choosing a metric in each class with zero Ricci scalar, and integrating over conformal factors and then over all equivalence classes. The first integral would converge if the contour of integration were rotated to lie parallel to the imaginary axis. However, this prescription is limited in that the requirement that $R = 0$ means that some metrics are missed out from the sum–over–geometries, and the rotation of the gravitational part of the path integral could destroy the positivity of a matter action which is not conformally invariant. These limitations are especially troublesome in the case of closed cosmologies. Here, there is also the further problem of there being no analogue to the positive action theorem which states that all Euclidean metrics which asymptotically approach flat space at infinity have positive action. Thus for closed cosmologies, the conformal equivalence class prescription fails to ensure a convergent integral.

A way to avoid these difficulties is to abandon the attempts to formulate a prescription for the contour within the preconceived framework of Euclidean gravity, but rather to place the emphasis on searching for all contours which produce consistent and physically reasonable results. Several constraints on the contour defining a wavefunction of the universe then suggest themselves. For instance, the requirement of a classical spacetime discussed in the previous chapter could be one such constraint. The most important constraint, however, is the necessity of a convergent integral, without which the wavefunction cannot even be defined. As has recently been discussed by several authors [16-18], the convergence requirement does not place many restrictions on the contours available, but it does imply that the integral must be over complex metrics. Then the extrema of the action which dominate the integral will have both real and imaginary parts:

$$I\,[g_0] = I_R[g_0] \pm iS[g_0]\,. \tag{2.1.7}$$

Generally, the whole action, I, will satisfy the Euclidean Hamilton–Jacobi equation and not either the real or imaginary parts on their own. However, when the three–metric, h, is large, the imaginary part of the action, S, will vary more rapidly with h than the real part, I_R. Thus at late times S alone will satisfy the Hamilton–Jacobi equation to a good approximation. This means that the wavefunction would then take an oscillatory form representing the classical, Lorentzian universe. The real part of the action, I_R, would enter only as an exponential damping term in the prefactor of the wavefunction.

In the language of tunnelling offered by gravitational instantons, the use of complex solutions provides a picture of a transition from a Euclidean universe to a Lorentzian one; the latter being predicted with a probability e^{-I_R}. The simplest example of such a picture is if the action describes pure gravity with a cosmological constant. Then a suitable solution to the Einstein equtions is that of a four–sphere joined across an equator to a Lorentzian de Sitter space at its minimal radius. This solution is a minimum of the real part of the action and therefore, like the other examples of instantons described above, dominates the transition amplitude. The properties of the join are derived from the matching condition of contiuity ofthe extrinsic curvature, K_{ij}, which is defined by

$$K_{ij} = \frac{1}{2N}\left[\frac{\partial h_{ij}}{\partial \tau} + D_{(i}N_{j)}\right]\,, \tag{2.1.8}$$

where D_i is the derivative in the three–surface. Since K_{ij} is purely real in Euclidean spacetimes and purely imaginary in Lorentzian ones, this implies that the metrics can only be joined across a surface with zero extrinsic curvature – that is, at the equator of the four–sphere and the throat of de Sitter space.

The above examples of tunnelling in cosmology illustrate applications of gravitational instantons. By analogy with the insantons of other field theories, gravitational instantons may be defined as non–singular finite action solutions of the Euclidean Einstein equations. The example of a universe tunnelling into de Sitter space uses instantons which are solutions of the Euclidean Einstein equations with a cosmological term Λ, so that Λ is part of the dynamics of the theory. However, Λ may also be introduced as a Lagrange multiplier for the four–volume V. This was the interpretation adopted by Hawking [48] in his analysis of the spacetime foam, which provides a further example of the applications of gravitational instantons.

The technique involves the evaluation of the partition function, Z. The partition function specifies the finite temperature behaviour of any theory and is defined by

$$Z = \mathrm{Tr}\,(e^{-\beta H})\,, \tag{2.1.9}$$

where $\beta = T^{-1}$ is the inverse temperature and H is the Hamiltonian operator of the system. For a gravitational system with a cosmological term, Z may be rewritten in the path integral formulation as

$$Z[\Lambda] = \int d[g]\, e^{-I[g]}\,, \tag{2.1.10}$$

where the integral is performed over all metrics g on a compact manifold M, and I denotes the Euclidean action of the gravitational fields with a Λ term. A semiclassical approximation implies that the dominant contributions to the integral will come from metrics near a solution, g_0, to the classical Einstein equations. Then the partition function will be approximately given by

$$Z \sim e^{-I[g_0]}\,. \tag{2.1.11}$$

There is a clear analogy with the vacuum–vacuum transition amplitude in the presence of a source J, (2.1.3). However, this physical interpretation cannot be applied here since the manifold is compact and thus there are no initial and final asymptotic regions.

The idea of spacetime foam was first introduced by Wheeler [49]. The lack of scale invariance for gravitational fields implies that at short scales of the order of the Planck length there may be very large fluctuations of the metric which have only small action. These contributions would not be highly damped and thus one arrives at a picture of a spacetime which is nearly flat at large length scales, but is highly curved on the scale of the Planck length. Hawking provided a mathematical framework for this picture by adding a source term $\Lambda V/8\pi$ to the usual gravitational action, $-1/16\pi \int R\sqrt{-g}\, d^4x$, where V is a unit four–volume. For high Euler number,* χ, the path integral is then dominated by solutions with action $-\chi\Lambda^{-1}$. This action can be interpreted as χ gravitational instantons, each of action $-\Lambda^{-1}$. Analysis of the one–loop terms implies that dominant contributions come from metrics with $\chi \sim V$, that is, with one instanton per unit Planck volume. Thus one finds a foam–like structure.

The topological invariants, the Euler number and the signature, can be expressed in terms of nuts and bolts - concepts developed by Gibbons and Hawking [50] to classify gravitational instantons. Nuts are fixed points of the Killing vector field and represent singularities of a polar co–ordinate system which can be removed by changing to a local Cartesian co–ordinate system near the point of singularity and adding that point to the manifold. The standard example of a nut is the self–dual Taub–NUT metric, after which the singularity is named. A bolt is a fixed surface of the Killing vector field and therefore represents singularities of the Rindler co–

* χ is given by

$$\chi = (128\pi^2)^{-1} \int_M R_{abcd} R_{efgh} \epsilon^{abef} \epsilon^{cdgh} \sqrt{-g}\, d^4x \ ,$$

and is one of two topological invariants which can be used to classify a simply connected manifold. The other invariant is the signature, τ, given by

$$\tau = (96\pi^2)^{-1} \int_M R_{abcd} R^{ab}{}_{ef} \epsilon^{cdef} \sqrt{-g}\, d^4x \ .$$

ordinate system in Minkowski space. Again, these may be removed by changing to local Cartesian co–ordinates. The canonical example of a bolt is the Schwarzschild metric.

A nut contributes one unit to the Euler characteristic of the manifold, whilst a bolt usually contributes two units. In this way, gravitational instantons can be classified; such a classification can be found in Eguchi, Gilkey and Hanson [51].

This section has sought to explain what instantons are, and, through the description of some of their applications, why they might be of interest. Attention has centred on gravitational instantons, and in particular those which are solutions of the Einstein equations with cosmological constant. This has been by way of motivation for the rest of the chapter in which new instantons are derived from the metrics of black holes in de Sitter space.

2.2 Black Hole Metrics

Since the advent of the general theory of relativity, the notion of a black hole has become widely understood - even to the extent that the term has entered common parlance, a rare occurance for theoretical concepts. A black hole forms after the gravitational collapse of a sufficiently massive body. The resulting spacetime contains a region from which no light ray can 'escape'; any observer entering this region will be trapped by the gravitational field and will eventually fall into a spacetime singularity. This singularity is contained within the black hole and cannot be seen by any observer remaining outside.

The simplest example of a black hole is to be found in the Schwarzschild metric, which is the only spherically symmetric vacuum solution of Einstein's equations and which describes the geometry exterior to a spherical collapsing body. It is of more general interest to look at the results of non–spherical collapse. It is reasonable to assume that the black hole formed by any such collapse should eventually settle down asymptotically in time to some stationary equilibrium state. The only vacuum solutions of Einstein's equations describing stationary black holes are the Kerr metrics. These are characterised by the parameters M, representing the mass, and a, the angular momentum. This class of solutions represents rotating black holes and contains the Schwarzschild solution as the special case with zero angular momentum. Generalisations of these metrics to solutions of the source free Einstein–Maxwell equations exist. As might be expected, these are described by a further two parameters: the electric charge, Q, and the magnetic monopole charge, P. The Reissner–Nordström metric is the charged generalisation of the Schwarzschild metric, whilst the remaining stationary electrovac solutions are known as the Kerr–Newman family of solutions.

The above metrics all describe spacetimes which are asymptotically flat, and it is only within such spacetimes that black holes can be properly defined. A curved spacetime can be thought of as asymptotically flat if the metric structure at null infinity is similar to that of Minkowski space. In this definition is an implicit assumption that we are able to describe the structure of spacetime at infinity. This is possible via a representation due to Penrose, in which a suitable transformation of co–ordinates enables one to rewrite the metric, $g_{\mu\nu}$, in the form

$$g_{\mu\nu} = \Omega^{-2}\tilde{g}_{\mu\nu} \,, \tag{2.2.1}$$

where Ω is a conformal factor and $\tilde{g}_{\mu\nu}$ is the metric on a new, unphysical manifold $\tilde{M}$. In the case of Minkowski space the original metric

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.2.2}$$

can yield a conformal metric

$$d\bar{s}^2 = -dT^2 + dR^2 + R^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.2.3}$$

valid in the range

$$-\pi < T + R < \pi \,, \quad -\pi < T - R < \pi \,, \quad R \geq 0 \,. \tag{2.2.4}$$

The metric (2.2.3) is identical to that of the Einstein static universe. Thus the whole of Minkowski spacetime is conformal to the region of the Einstein static universe specified by conditions (2.2.4). The boundary of this region represents the conformal structure of infinity of Minkowski spacetime, and consists of the null surfaces $\Im^+$ (future null infinity) and $\Im^-$ (past null infinity) and the points i^+ (future timelike infinity), i^- (past timelike infinity) and i^0 (spacelike infinity). Null geodesics originate at $\Im^-$ and end at $\Im^+$; timelike geodesics originate at i^- and end at i^+; whilst spacelike geodesics both start and finish on i^0. This conformal structure of infinity is most clearly represented by a diagram of the (T,R) plane (fig.2.1). Here infinity is represented by single lines, the origin of the polar co–ordinates by a broken line. Each point of the diagram represents a two–sphere, and null geodesics are shown as straight lines at $\pm 45^o$. This type of diagram is known as a Penrose diagram.

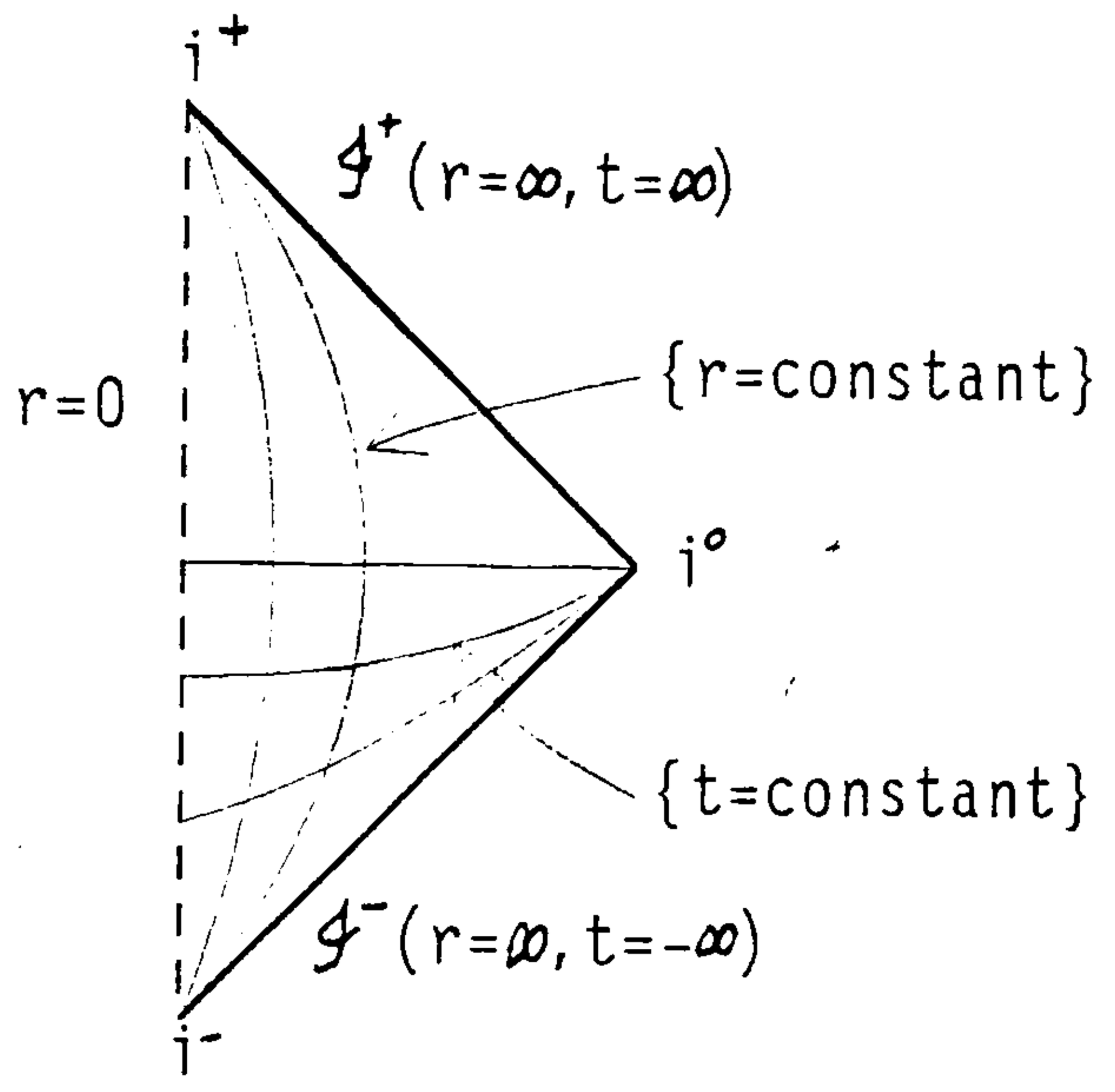

Fig.2.1 Penrose diagram of Minkowski spacetime.

In a similar manner, a representation of the structure of infinity in a more general spacetime, $(M, g_{\mu\nu})$, can be arrived at by embedding M as a manifold with smooth boundary ∂M in another manifold $\tilde{M}$, which has metric $\tilde{g}_{\mu\nu}$ conformal to $g_{\mu\nu}$. If such a mapping is possible and produces a structure similar to that obtained for Minkowski spacetime, then the spacetime $(M, g_{\mu\ \nu})$ is said to be asymptotically flat.

Given an asymptotically flat spacetime, the gravitational collapse of any body sufficiently close to having spherical symmetry must result in a singularity. However, the formation of a black hole requires more stringent properties of the spacetime. This involves the notion of strong future asymptotic predictability. In essence, a strongly asymptotically predictable spacetime is one in which all events visible from infinity can be predicted from conditions at an instant of time represented by a partial

Cauchy surface, $\mathcal{J}$. Such a surface is a spacelike hypersurface which no non–spacelike curve intersects more than once. The future Cauchy development of $\mathcal{J}$ in a manifold M is a region $D^+(\mathcal{J})$ containing all points $p \in M$, such that every past–inextendible non–spacelike curve through p intersects $\mathcal{J}$. Thus $D^+(\mathcal{J})$ is the region of the universe which can be predicted by knowledge of the appropriate data on $\mathcal{J}$, since all points outside this region, being joined to $\mathcal{J}$ by spacelike curves, are not causally connected to $\mathcal{J}$. A strongly asymptotically predictable spacetime can be defined as having a partial Cauchy surface, $\mathcal{J}$, such that future null infinity, $\Im^+$, lies in the boundary of $D^+(\mathcal{J})$, and $J^+(\mathcal{J}) \cap \bar{J}^-(\Im^+)$ is contained within $D^+(\mathcal{J})$, where $J^+(\mathcal{J})$ is the causal future of $\mathcal{J}$, and $\bar{J}^-(\Im^+)$ is the closure of the causal past of $\Im^+$.

In a strongly asymptotically predictable spacetime, one can give a definition of a black hole region, B, as being that region lying outside the causal past of future null infinity $\Im^+$. That is, for a manifold M,

$$B = [M - J^-(\Im^+)] \, . \tag{2.2.5}$$

All non–spacelike curves through points in B fail to reach $\Im^+$ but end in the singularity contained within B. The event horizon, H, is the boundary of the black hole region in M,

$$H = M \cap \dot{J}^-(\Im^+) \, . \tag{2.2.6}$$

The condition of asymptotic predictability implies that no singularity is visible to any observer in $D^+(\mathcal{J})$, since a visible singularity would have future directed non–spacelike curves which ended on $\Im^+$ but which were inextendible in the past and therefore failed to intersect the surface $\mathcal{J}$. Conversely, a spacetime which fails to be strongly asymptotically predictable will possess a 'naked' singularity. It is believed that all physically relevant spacetimes are strongly asymptotically predictable. This leads to the cosmic censorship conjecture in which it is asserted that the complete gravitational collapse of a body always results in a black hole rather than a naked singularity. Thus all singularities, with the possible exception of an initial singularity, are hidden from the observer within black holes.

The above definitions of a black hole and event horizon applied only to asymptotically flat spacetimes where one can provide a clear representation of infinity and thus formalise the idea of light not being able to 'escape to infinity'. However, the metrics to be studied in this chapter describe black holes in de Sitter space. Even in empty de Sitter space, there already exists an event horizon. By a construction similar to that followed above for Minkowski space, one may draw a Penrose diagram of de Sitter space (fig.2.2).

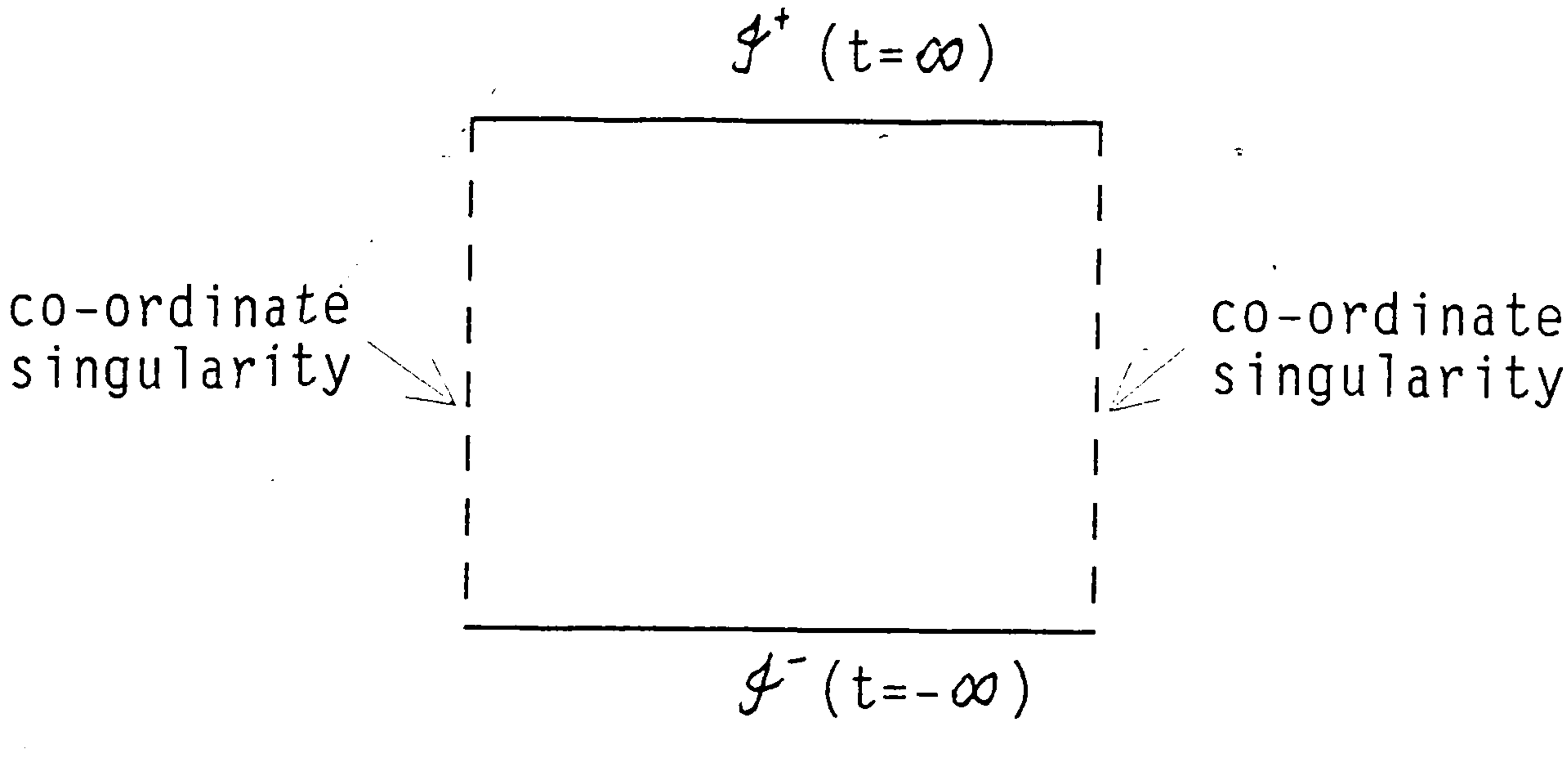

Fig.2.2 Penrose diagram of de Sitter space.

The important feature of this diagram is that future and past infinity for null and timelike curves are spacelike. This means that there are some events which will never be observable by an observer. The world line of the observer must end somewhere on $\mathcal{I}^+$, but only events lying within the past null cone of this point can ever be seen by the observer (fig.2.3). The boundary marked by this past null cone is the future event horizon of the spacetime.

The absence of a future null infinity also implies that the definition of a black hole as that region from which light rays fail to reach null infinity is inadequate.

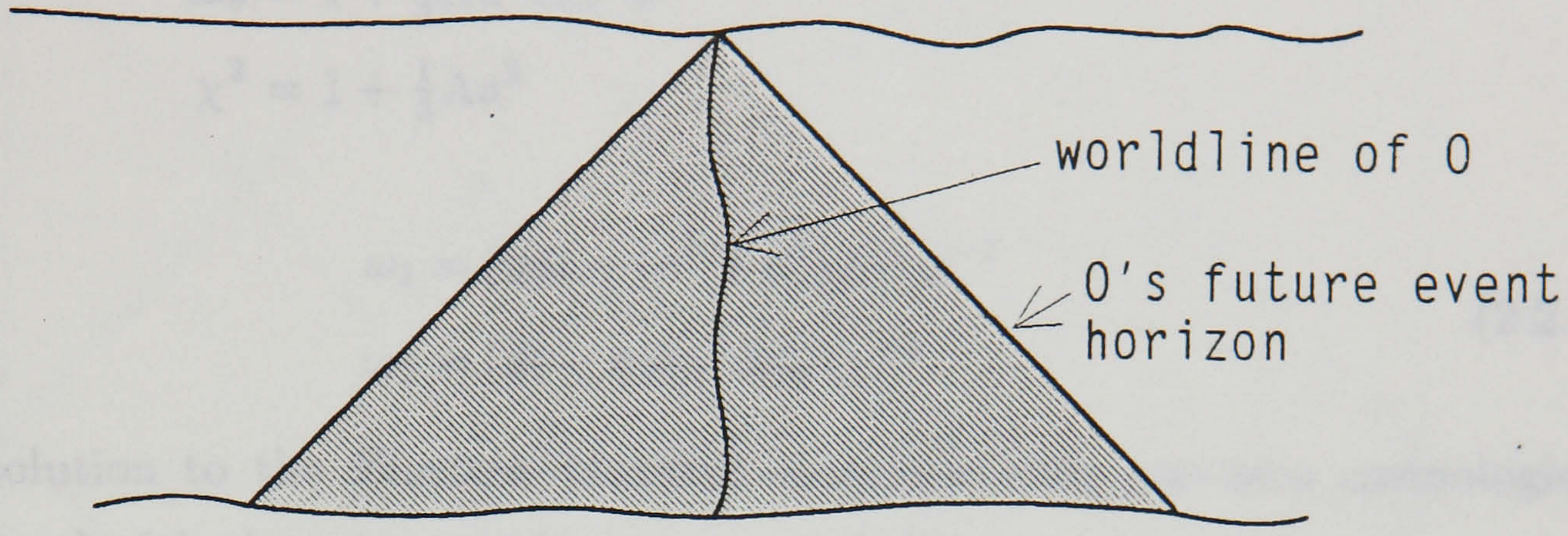

Fig.2.3 Event horizon due to spacelike nature of $\Im^+$.

However, the notion of a black hole in de Sitter space can be retrieved since the spacelike surfaces at infinity do carry some sense of an asymptotic region. All timelike geodesics with an endpoint terminate on the future infinity and so the past null cone of this surface represents that part of the universe which will be visible from infinity. In this way, the definition of a black hole event horizon as the boundary of the causal past of infinity may be carried over for a given asymptotic region of de Sitter space.

Thus it is possible to define in de Sitter space a charged black hole rotating with angular momentum a. The general metric for such a situation has been given by Carter [52] and takes the form

$$ds^2 = \rho^2(\Delta^{-1}dr^2 + \Delta_\theta^{-1}d\theta^2) + \rho^{-2}\Delta_\theta \sin^2\theta\omega_1^2 - \rho^{-2}\Delta\omega_2^2 \qquad (2.2.7)$$

where

$$\begin{aligned}
\rho^2 &= r^2 + a^2 \cos^2 \theta \\
\Delta &= (a^2 + r^2)(1 - \tfrac{1}{3}\Lambda r^2) - 2Mr + Q^2 + P^2 \\
\Delta_\theta &= 1 + \tfrac{1}{3}\Lambda a^2 \cos^2 \theta \\
\chi^2 &= 1 + \tfrac{1}{3}\Lambda a^2
\end{aligned} \tag{2.2.8}$$

and

$$\begin{aligned}
\omega_1 &= (a dt - (r^2 + a^2) d\phi)\, \chi^{-2} \\
\omega_2 &= (dt - a \sin^2 \theta d\phi)\, \chi^{-2}
\end{aligned} \tag{2.2.9}$$

This is a solution to the Einstein–Maxwell equations with non-zero cosmological constant Λ and with electromagnetic vector potential

$$A = \frac{P \cos \theta}{\rho^2}\, \omega_1 + \frac{Qr}{\rho^2}\, \omega_2 \tag{2.2.10}$$

corresponding to electric and magnetic charges Q and P.

The metric (2.2.7) has co–ordinate singularities at the roots of $\Delta = 0$. The presence of a cosmological constant makes this a quartic equation, so that for a range of parameters there are four real roots. These shall be labelled, in decreasing value of r, as r_1, r_2, r_3, r_4. The fourth root is negative and therefore non–physical. The singularities at the other roots are features of the co–ordinate system alone and may be removed by suitable co–ordinate transformations in the appropriate regions. This will be demonstrated in §2.3 below. The roots then represent horizons on which the metric is regular.

The horizon at r_1 is a de Sitter event horizon, whilst r_2 and r_3 represent the outer and inner horizons respectively of the black hole. The outer horizon is the usual black hole event horizon which marks the boundary of events visible from infinity. However, there is a further horizon at r_3. Beyond r_3 there will exist past directed inextendible timelike curves which approach an irremovable singularity rather than the asymptotic region outside the black hole. Thus, if one imagines a Cauchy surface, $\mathcal{J}$, in the asymptotic region, there is a boundary to the future Cauchy development $D^+(\mathcal{J})$ of $\mathcal{J}$. This boundary is the inner horizon of the black hole. Events beyond this horizon cannot be predicted by data on $\mathcal{J}$.

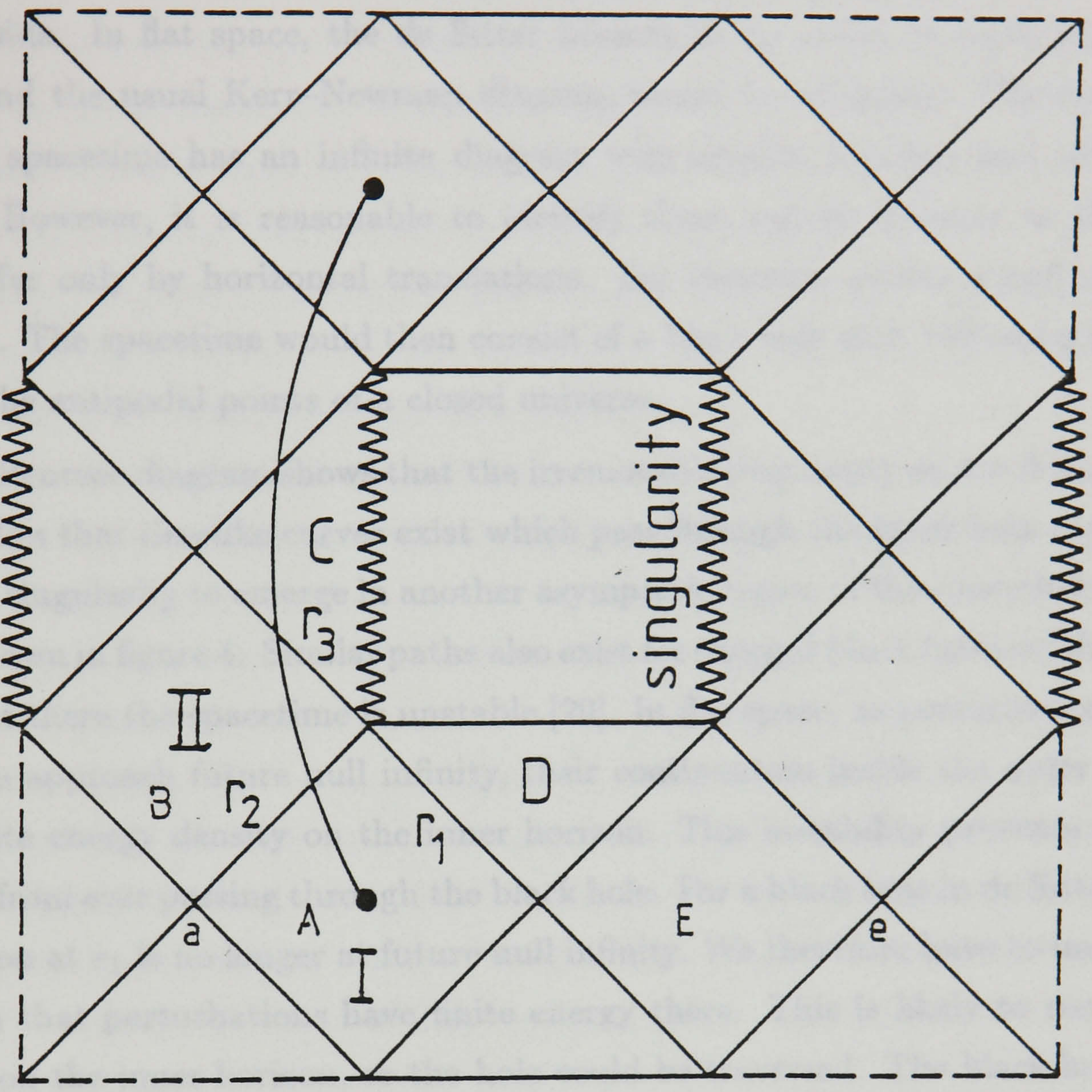

Fig. 2.4

Penrose diagram of charged and/or rotating black hole in de Sitter space.

The situation is best illustrated by a Penrose diagram (fig.2.4). In this figure, the horizons are labelled r_1, r_2, r_3 as above, and the irremovable singularity is represented by a wavy line. Regions B and C lie inside the black hole, while regions A,D and E lie outside. In flat space, the de Sitter horizon at r_1 would be replaced by null infinity and the usual Kerr–Newman diagram would be obtained. The maximally extended spacetime has an infinite diagram with infinite horizons and black hole regions. However, it is reasonable to identify those regions exterior to the holes which differ only by horizontal translations. For instance, points a and e can be identified. The spacetime would then consist of a black hole with two outer horizons lying at the antipodal points of a closed universe.

The Penrose diagram shows that the irremovable singularity at $\rho = 0$ is timelike. This implies that timelike curves exist which pass through the black hole region, but avoid the singularity to emerge in another asymptotic region of the spacetime. Such a path is shown in figure 4. Similar paths also exist for charged black holes in Minkowski space, but there the spacetime is unstable [20]. In flat space, as perturbations of the spacetime approach future null infinity, their continuation inside the outer horizon has infinite energy density on the inner horizon. This instability prevents any real observer from ever passing through the black hole. For a black hole in de Sitter space, the horizon at r_1 is no longer at future null infinity. We therefore have to impose the condition that perturbations have finite energy there. This is likely to remove the problem on the inner horizon, so the hole could be traversed. The black hole could then be thought of as a wormhole or spacetime 'bridge' which joins two asymptotic regions of the universe. The stability of the spacetime is the subject of the following chapter.

It should also be noted that, as in the case of the Kerr solution in flat space, the Killing vectors of metric (2.2.7) have interesting properties. In the metrics of Schwarzschild or Reissner–Nordström,

$$ds^2 = -r^{-2}\Delta dt^2 + r^2\Delta^{-1}dr^2 + r^2(d\theta^2 + sin^2\theta d\phi^2) \qquad (2.2.11)$$

where

$$\Delta = r^2 - 2Mr + Q^2 + P^2 \ , \qquad (2.2.12)$$

there are two Killing vectors given by the basis vectors of the co–ordinate system: $\partial/\partial t$ and $\partial/\partial\phi$. A Killing vector indicates that the metric is invariant under flow along that vector field and thus describes the symmetries of the metric. The Killing vector $\partial/\partial\phi$ implies that metric (2.2.11) is axisymmetric, whilst $\partial/\partial t$ indicates the stationary nature of the metric. Thus the orbits of $\partial/\partial t$ are the curves along which a particle must travel in order to appear at rest with respect to an observer at infinity.

Metric (2.2.11) has two horizons r_+ and r_-: r_+ represents the outer black hole event horizon; r_- is the inner horizon. For values of r greater than r_+, that is, outside the black hole region, $\partial/\partial t$ is timelike. However, on the horizon r_+, the Killing vector becomes null, and within the black hole region, $r_+ > r > r_-$, it is spacelike. This means that in this region it is impossible for a particle travelling along a timelike or null curve to be following the orbit of a Killing vector and thus to be at rest with respect to infinity.

In the case of a Kerr solution, whether in flat spacetime or de Sitter, there is again a two–parameter group of isometries reflecting the stationary and axisymmetric nature of the solution. However, the surface on which the Killing vector $\partial/\partial t$ becomes null no longer coincides with the outer black hole horizon. Instead, there is a region in which the Killing vector is spacelike which lies outside the black hole. The boundary of this region, the surface on which $\partial/\partial t$ goes null, is known as the stationary limit surface, and the region defined by this surface and the outer black hole is called the ergosphere (fig.2.5).

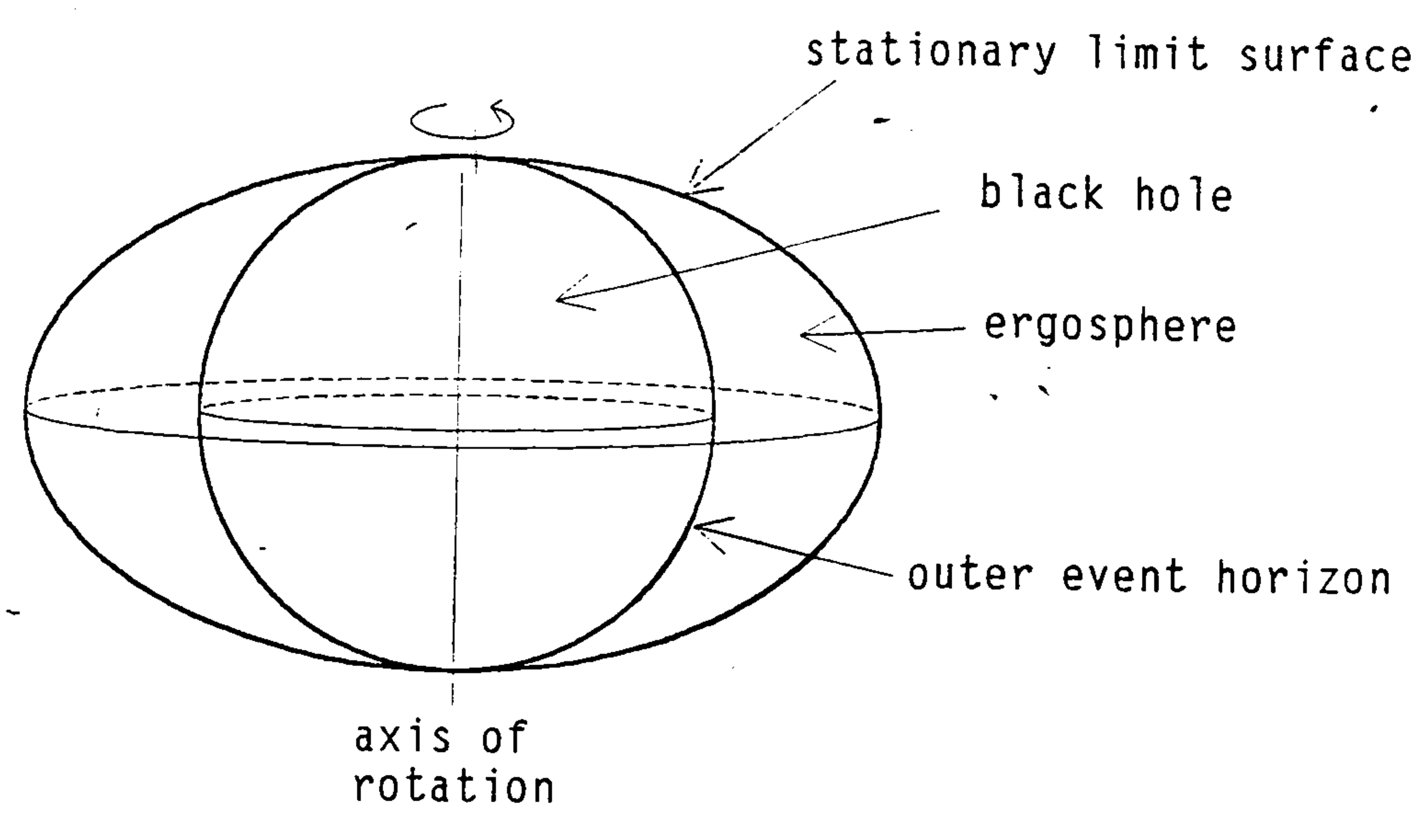

Fig.2.5 The ergosphere of a Kerr solution.

Although the behaviour of the Killing vector $\partial/\partial t$ inside the ergosphere implies that observers in this region cannot be stationary with respect to infinity, the solutions here may be thought of as locally stationary. This is because it is possible to construct a Killing vector, l, which is a linear combination of the Killing vectors $\partial/\partial t$ and $\partial/\partial\phi$, and which is null on the event horizon and timelike outside it. The appropriate combination is

$$l = \frac{\partial}{\partial t} + \Omega_H \frac{\partial}{\partial \phi}\,, \tag{2.2.13}$$

where Ω_H is the rotational velocity of the horizon and is given by

$$\Omega_H = \frac{a}{r_H^2 + a^2}\,. \tag{2.2.14}$$

The existence of a Killing vector which coincides with the null generators of the horizon enables a quantity, κ_i, to be defined which is given by

$$\kappa_i^2 = \lim_{r \to r_i} \frac{1}{4}\frac{\left(\nabla|l^2|\right)^2}{|l^2|}\,, \tag{2.2.15}$$

where r_i denotes the appropriate horizon. κ_i is constant on the horizon and in the case of a static black hole can be given the physical interpretation of surface gravity. That is, it is the limiting value at the horizon of the force required to hold a unit test mass in place. Although this interpretation can no longer apply for a rotating black hole since it is then impossible to hold a particle stationary near the black hole, κ_i is still referred to as the surface gravity.

The general black hole in de Sitter space, with metric (2.2.7) and Killing vector (2.2.13), has surface gravity

$$\kappa_i = \tfrac{1}{6}\Lambda\chi^{-2}(r_i^2 + a^2)^{-1}\prod_{i \neq j}|r_i - r_j|\,. \tag{2.2.16}$$

The surface gravity plays an important role in black hole physics. The fact that it is constant over the horizon of a stationary black hole is suggestive of a body in thermal equilibrium having a constant temperature throughout. The thermodynamic parallels are strengthened by the area theorem of black holes which states that the area of any connected two–surface in an event horizon cannot decrease with time.

Thus the area of a black hole may be compared to the entropy of a thermodynamic system, and the properties of the surface gravity and of the area may be thought of as the zeroth and second laws respectively of black hole thermodynamics. The analogy can be extended to the first and third laws as well, but classically the comparisons between the two types of system do not appear to have any physical foundation and fail to go beyond mere analogy. This is because classical general relativity describes a black hole as a perfect absorber of radiation. Since this implies that the hole emits no particles, the temperature of the hole must be absolute zero. Thus the surface gravity cannot represent the physical temperature. In the quantum regime, however, a black hole does emit particles [12]. The mechanism by which this occurs can be understood as pair creation in the gravitational field of the the black hole. One particle, with positive energy, will escape to infinity and this contributes to the black hole radiation. The other particle, having negative energy, falls into the hole where the spacelike nature of the Killing vector implies that negative energy states can exist. Such a process produces particles at the same rate as is expected from a body at temperature $T = \kappa/2\pi$. In doing so the hole loses mass and therefore heats up. Thus quantum effects suggest that it is legitimate to go beyond analogy and to assign to a multiple of the surface gravity the physical significance of a temperature.

The laws of black hole thermodynamics and the phenomenon of Hawking radiation also apply to cosmological horizons [53]. Thus the event horizon of de Sitter space emits thermal radiation at a temperature corresponding to the surface gravity on that horizon. It follows that a spacetime consisting of both a black hole horizon and a non–zero cosmological constant, like (2.2.7), will have two sources of thermal radiation. If the temperatures of these sources are different, then energy flows from the hotter body to the cooler and the black hole will gain or lose mass. For a stable spacetime we therefore need to find the conditions under which the temperatures are equal.

The difference between the surface gravities of the outer black hole horizon and the de Sitter horizon of (2.2.7) can be manipulated into the form

$$\kappa_1 - \kappa_2 = -\frac{\Lambda}{6}\frac{\chi^{-2}(r_1 + r_2)}{(r_1^2 + a^2)(r_2^2 + a^2)}(r_1 - r_2)^2(r_1 r_2 + r_3 r_4 - 2a^2)\ , \tag{2.2.17}$$

One obvious family of solutions with $\kappa_1 = \kappa_2$ are those with coincident horizons at $r_1 = r_2$. In these cases it follows from (2.2.16) that $\kappa = 0$. Although these solutions are in thermodynamic equilibrium, they are unstable to changes in the mass of the black hole. If the mass decreases, then $\kappa_1 - \kappa_2$ is negative because the last factor in (2.2.17) is positive. Thus energy will flow from the de Sitter horizon to the outer black hole horizon causing the hole to become hotter and to continue to shrink.

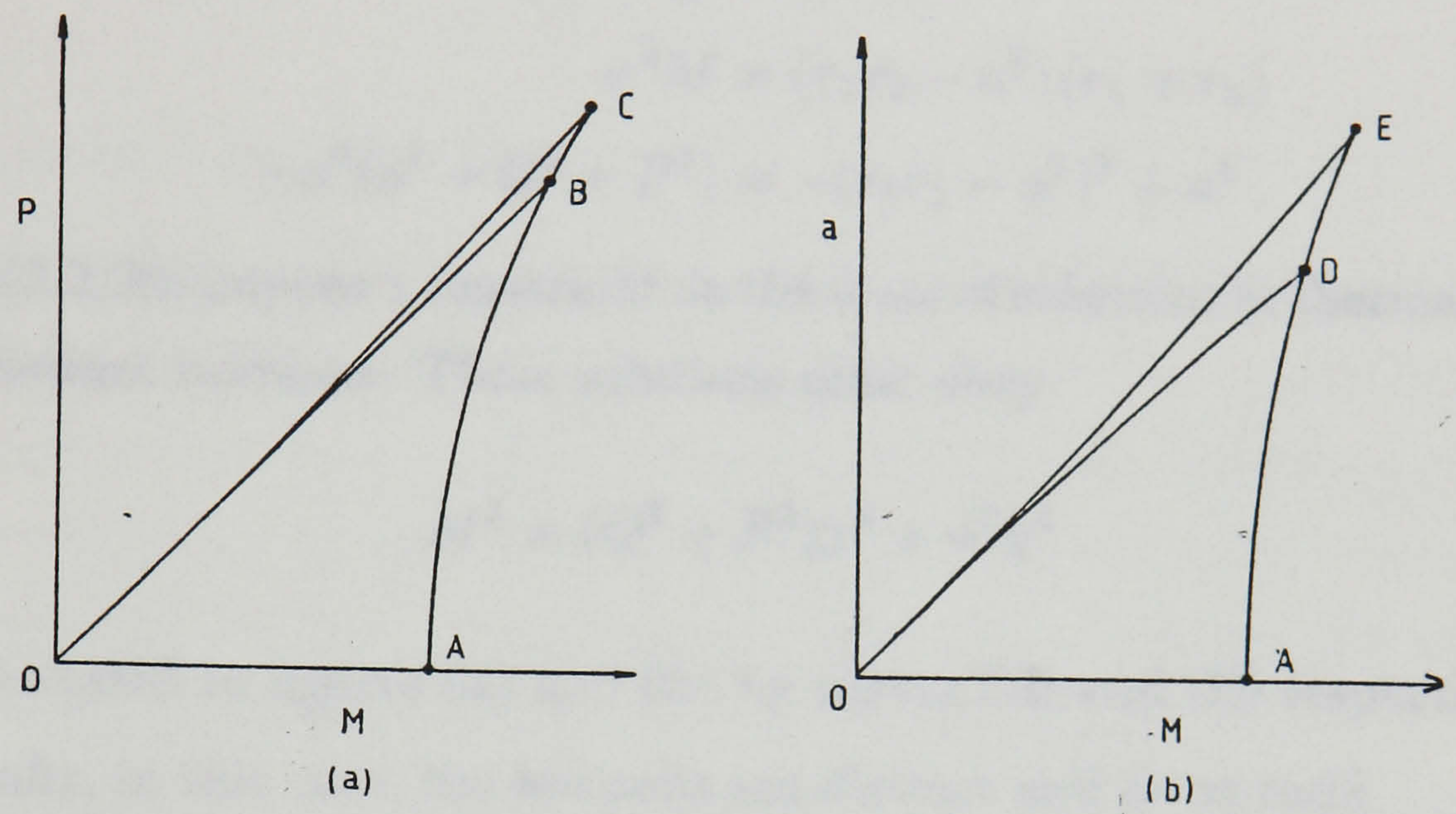

Fig.2.6 The parameter space of solutions with equal surface gravity.

Some sections of the parameter space of solutions are shown in figure 2.6. In figure 6a) the rotation parameter a has been set to zero. The curve AC shows the solutions described above for which the black hole and de Sitter horizons coincide. Also shown is the condition $r_2 = r_3$, where the inner and outer black hole horizons coincide. These solutions lie along the curve OC. Thus the region for which there exists three distinct horizons lies within OAC. Figure 6b) shows the corresponding curves with the charge Q set to zero rather than the rotation parameter.

The instability of the thermodynamic equilibrium for solutions with coincident horizons at r_1 and r_2 makes these solutions of little interest. A more significant family of solutions with $\kappa_1 = \kappa_2$ are those for which the last factor in (2.2.17) vanishes; i.e. those solutions satisfying the condition

$$(r_1 r_2 + r_3 r_4 - 2a^2) = 0\ . \tag{2.2.18}$$

The horizons are the roots of the factor Δ in (2.2.8), and are therefore given by

$$-\alpha^{-2}\{r^4+(a^2-\alpha^2)r^2+2\alpha^2Mr-\alpha^2(a^2+Q^2+P^2)\} = -\alpha^{-2}\prod_{i=1\text{ to }4}(r-r_i) \quad (2.2.19)$$

where $\alpha^2 = 3/\Lambda$. Equating coefficients in (2.2.19), together with condition (2.2.18) gives

$$\begin{aligned}(r_1^2+r_2^2) &= a^2+\alpha^2 \\ \alpha^2 M &= (r_1r_2-a^2)(r_1+r_2) \\ -\alpha^2(a^2+Q^2+P^2) &= -(r_1r_2-a^2)^2+a^4 \,.\end{aligned} \quad (2.2.20)$$

Equations (2.2.20) impose a constraint on the mass of solutions in thermal equilibrium but with distinct horizons. These solutions must obey

$$M^2 = (Q^2+P^2)\chi^2 + a^2\chi^4 \,, \quad (2.2.21)$$

which is illustated in figures 6a) and 6b) by curves OB and OD respectively.

Generally, in this case, the horizons are distinct and lie at radii

$$\begin{aligned} r_1 &= \tfrac{1}{2}\alpha\chi + \tfrac{1}{2}(\alpha^2\chi^2-4a^2-4\alpha M\chi^{-1})^{1/2} \\ r_2 &= \tfrac{1}{2}\alpha\chi - \tfrac{1}{2}(\alpha^2\chi^2-4a^2-4\alpha M\chi^{-1})^{1/2} \\ r_3 &= -\tfrac{1}{2}\alpha\chi + \tfrac{1}{2}(\alpha^2\chi^2-4a^2+4\alpha M\chi^{-1})^{1/2} \end{aligned} \quad . \quad (2.2.22)$$

On both horizons r_1 and r_2 the surface gravity is given by

$$\kappa = \alpha^{-1}\chi^{-2}(1-4M/\alpha\chi-3a^2/\alpha^2)^{1/2} \,. \quad (2.2.23)$$

The thermodynamic considerations discussed so far have ignored the possibility that the black holes radiate particles by non-thermal processes. However, rotating holes can radiate into superradiant wave modes [54] and lose some angular momentum. These modes have reflection coeficients which are greater than unity. Each time the radiation is reflected near the de Sitter or black hole horizons it increases in amplitude and all the rotational energy is released catastrophically.

Non-rotating holes can also radiate charge [55], causing the holes to discharge and evaporate away. However, this may be prevented if the charge on the hole is not the ordinary electric one and if there are no suitably charged particles which are lighter than the hole. In this latter case, the black hole produces thermal radiation only. If $M^2 > Q^2 + P^2$, then one finds that $r_1 r_2 + r_3 r_4 > 0$ and therefore $\kappa_2 > \kappa_1$. Energy flows from the hole, which is hotter, and the mass of the hole decreases. Conversely, if $M^2 < Q^2 + P^2$, the mass increases. Thus the non-rotating version of condition (2.2.21), $M^2 = Q^2 + P^2$, represents a stable point in the spacetime.

2.3 Analytic Continuation

Euclidean instantons may be obtained from metrics of the form (2.2.7) by analytic continuation. This involves the complexification of the time co-ordinate through the introduction of a new co-ordinate $\tau = it$. The original metric components remain real because they are independent of t, but on the new manifold it is also necessary to complexify the angular momentum parameter by $\hat{a} = ia$ and the electric charge by $\hat{Q} = iQ$ because they are associated with single factors of dt.

Problems can arise with the Euclidean section due to a failure of the regularity of the space around the points r_1 and r_2. However, regularity can be demonstrated through a series of co-ordinate transformations. Consider first the horizon r_2. Initially transform to co-ordinates analogous to the Eddington–Finkelstein co-ordinates. That is, introduce an outgoing null co-ordinate $u = t - r^*$ and ingoing null co-ordinate $v = t + r^*$ where r^* is given by

$$dr^* = \chi^2 \Delta^{-1} (r^2 + a^2)\, dr \,. \tag{2.3.1}$$

Additionally, since for rotating black holes the null Killing vector, l, on the horizon is given by (2.2.13), a new azimuthal co-ordinate,

$$\hat{\phi} = \phi - \frac{a}{r_2^2 + a^2}\, t \,, \tag{2.3.2}$$

must be introduced, which is constant along the integral curves of l. Because l is a regular vector field on the horizon, $\hat{\phi}$ is expected to be a good co-ordinate there.

This transformation is succeeded by another to Kruskal co-ordinates

$$\begin{aligned} V &= e^{\kappa v} \\ U &= -\, e^{-\kappa u} \end{aligned} \tag{2.3.3}$$

with κ the surface gravity evaluated on horizon r_2. In these co–ordinates the metric takes the form

$$\begin{aligned} ds^2 =& -\frac{1}{2}\rho^2\frac{\Delta\chi^{-4}}{(r^2+a^2)^2}\kappa^{-2}e^{-2\kappa r^*}dUdV + \frac{1}{4}\rho^2\frac{\Delta\chi^{-4}}{(r^2+a^2)^2}\kappa^{-2}\left(U^{-2}dU^2 - V^{-2}dV^2\right) \\ &+\rho^2\Delta_\theta^{-1}d\theta^2 + \rho^{-2}\Delta_\theta\sin^2\theta\omega_1^2 - \rho^{-2}\Delta\omega_2^2 \end{aligned} \tag{2.3.4}$$

where

$$\begin{aligned} \chi^2\omega_1 =& \tfrac{1}{2}a\,(r_2^2-r^2)\,(r_2^2+a^2)^{-1}\kappa^{-1}(-U^{-1}dU+V^{-1}dV) - (r^2+a^2)d\hat{\phi} \\ \chi^2\omega_2 =& \tfrac{1}{2}(r_2^2+a^2\cos^2\theta)\,(r_2^2+a^2)^{-1}\kappa^{-1}(-U^{-1}dU+V^{-1}dV) - a\sin^2\theta d\hat{\phi}\,. \end{aligned} \tag{2.3.5}$$

To study the Euclidean section, the co–ordinates are complexified by

$$\begin{aligned} U =& -Re^{i\Theta} \\ V =& Re^{-i\Theta}\,. \end{aligned} \tag{2.3.6}$$

The metric then becomes

$$\begin{aligned} ds^2 =& \rho^2\chi^{-4}\Delta(r^2-\hat{a}^2)^{-2}\kappa^{-2}e^{-2\kappa r^*}(dR^2+R^2d\Theta^2) - \rho^2\chi^{-4}\Delta(r^2-\hat{a}^2)^{-2}\kappa^{-2}d\Theta^2 \\ &+\rho^2\Delta_\theta^{-1}d\theta^2 + \rho^{-2}\Delta_\theta\sin^2\theta\hat{\omega}_1^2 + \rho^{-2}\Delta\hat{\omega}_2^2 \end{aligned} \tag{2.3.7}$$

where

$$\begin{aligned} \chi^2\hat{\omega}_1 =& \hat{a}\,(r_2^2-r^2)\,(r_2^2-\hat{a}^2)^{-1}\kappa^{-1}d\Theta - (r^2-\hat{a}^2)d\hat{\phi} \\ \chi^2\hat{\omega}_2 =& (r_2^2-\hat{a}^2\cos^2\theta)\,(r_2^2-\hat{a}^2)^{-1}\kappa^{-1}d\Theta + \hat{a}\sin^2\theta d\hat{\phi}\,. \end{aligned} \tag{2.3.8}$$

The range $0 \le R < \infty$ covers the region $r_2 \le r < r_1$. Although the factor Δ in (2.3.7) goes to zero at r_2, the metric remains regular here. This is due to the factor $e^{-2\kappa r^*}$, whose r dependence can be seen from the explicit expression for r^*,

$$\begin{aligned} r^* = -\alpha^2\chi^2\Bigg\{ & \frac{\left(\frac{1}{2}(r-r_1)^2 + 2rr_1 + (r_1^2+a^2)\ln|r-r_1|\right)}{(r_1-r_2)(r_1-r_3)(r_1-r_4)} \\ &+\frac{\left(\frac{1}{2}(r-r_2)^2 + 2rr_2 + (r_2^2+a^2)\ln|r-r_2|\right)}{(r_2-r_1)(r_2-r_3)(r_2-r_4)} \\ &+\frac{\left(\frac{1}{2}(r-r_3)^2 + 2rr_3 + (r_3^2+a^2)\ln|r-r_3|\right)}{(r_3-r_1)(r_3-r_2)(r_3-r_4)} \\ &+\frac{\left(\frac{1}{2}(r-r_4)^2 + 2rr_4 + (r_4^2+a^2)\ln|r-r_4|\right)}{(r_4-r_1)(r_4-r_2)(r_4-r_3)}\Bigg\}\,. \end{aligned} \tag{2.3.9}$$

Thus, as r approaches r_2, $-2\kappa r^*$ varies as

$$-2\kappa r^* \sim \frac{2\kappa\alpha^2\chi^2(r_2^2+a^2)}{(r_2-r_1)(r_2-r_3)(r_2-r_4)}\ln|r-r_2|\,, \tag{2.3.10}$$

which, substituting for the surface gravity evaluated at r_2, yields

$$e^{-2\kappa r^*} \sim |r-r_2|^{-1}\,. \tag{2.3.11}$$

Since $\Delta \sim (r-r_2)$, the factors cancel and the metric remains regular as $r \to r_2$ and is given by

$$\begin{aligned} ds^2 =& \rho_2^2\chi^{-4}\Delta(r^2-\hat{a}^2)^{-2}\kappa^{-2}e^{-2\kappa r^*}(dR^2+R^2d\Theta^2) \\ &+ \rho_2^2\Delta_\theta^{-1}d\theta^2 - \rho_2^{-2}\chi^{-4}\Delta_\theta\sin^2\theta\left(r_2^2-\hat{a}^2\right)d\hat{\phi}^2\,, \end{aligned} \tag{2.3.12}$$

where ρ_2 is ρ evaluated at r_2. There is therefore only a trivial co–ordinate singularity at r_2, and Θ is interpreted as a polar angle. Since Θ is related to the original time co–ordinate by $\Theta = i\kappa t$, this implies that the Euclidean time must be periodic, with period $2\pi/\kappa$.

A similar metric can be constructed which is regular at r_1, by introducing co–ordinates

$$\begin{aligned} U' =& -e^{-\kappa(t-r^*)} = -R'e^{i\Theta'} \\ V' =& e^{\kappa(t+r^*)} = R'e^{-i\Theta'}\,. \end{aligned} \tag{2.3.13}$$

After analytic continuation, the metric obtained with these co–ordinates may be related to the earlier metric (2.3.7), by the co–ordinate transformation

$$R' = R^{-1} \tag{2.3.14}$$

providing that $|\kappa|$ is the same on both the horizons. This is the same as the condition for thermodynamic equilibrium discussed earlier. If this condition is met, then the Euclidean section can be covered by two regular co–ordinate patches as shown in figure 2.7.

The analytic continuation can be conveniently summarised in the terminology of nuts and bolts described in section §2.1. Originally, the region $r_2 < r < r_1$ lay between two bolts (fixed surfaces of a Killing field). Under analytic continuation these turn into nuts (fixed points of an angular Killing field). The Euler number

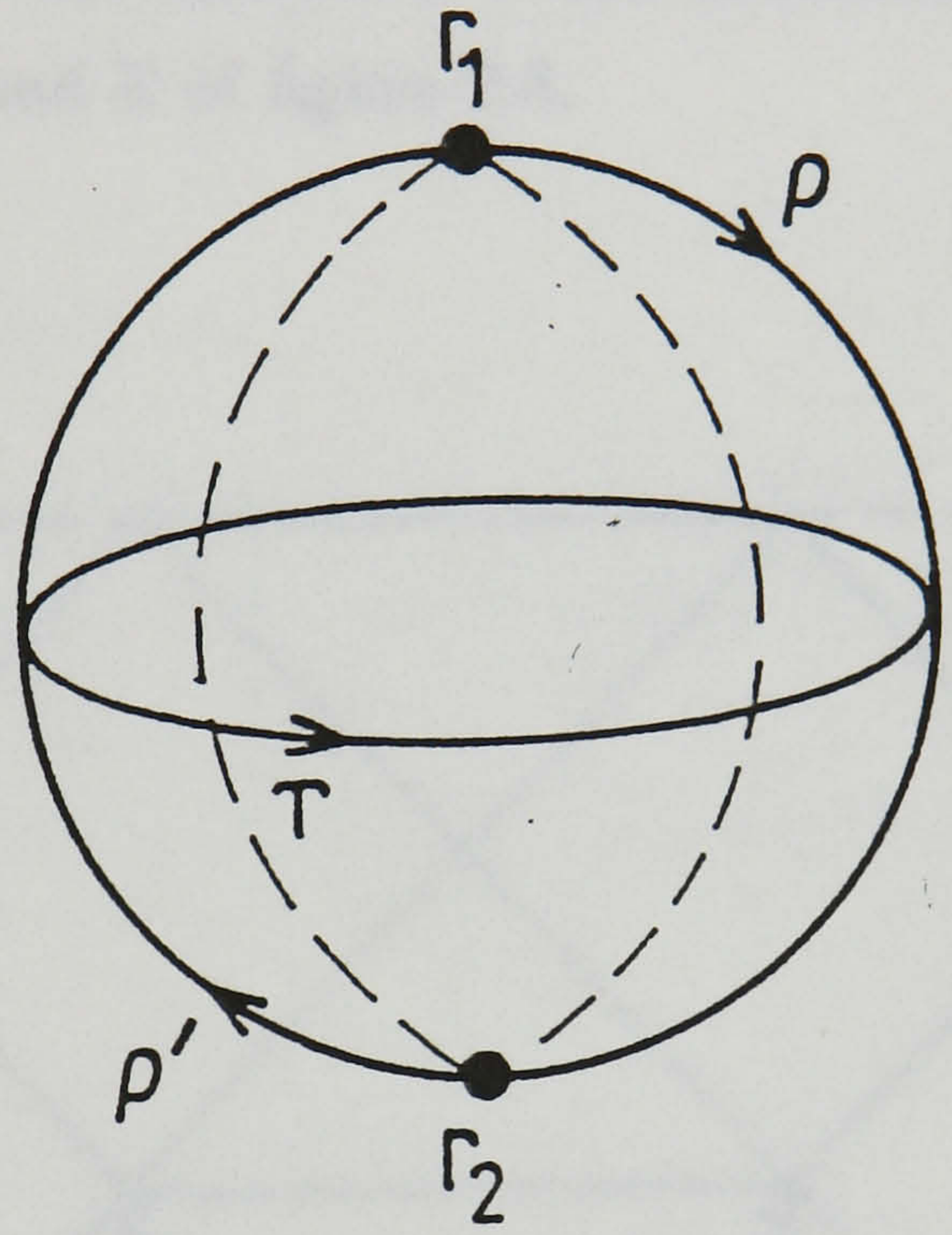

Fig 2.7 Schematic diagram of the Euclidean section. The broken curve shows a slice where $\tau = 0$ and $\tau = \pi$.

of the region equals the number of nuts and bolts plus two for the angles θ and ϕ. Consequently, the total space is topologically $S_2 \times S_2$.

The Euclidean section covers the region between the black hole and de Sitter space horizons. Slices of the metric correspond to great circles in figure 2.7 and are composed of two values of τ, for instance $\tau = 0$ and $\tau = \pi$. Slices such as this can be identified with time symmetric initial data for the Lorentzian spacetime. In particular, the geometry of the slice at $\tau = 0, \pi$ indicates that it represents initial data at $t = 0$ in regions A and E of figure 2.8.

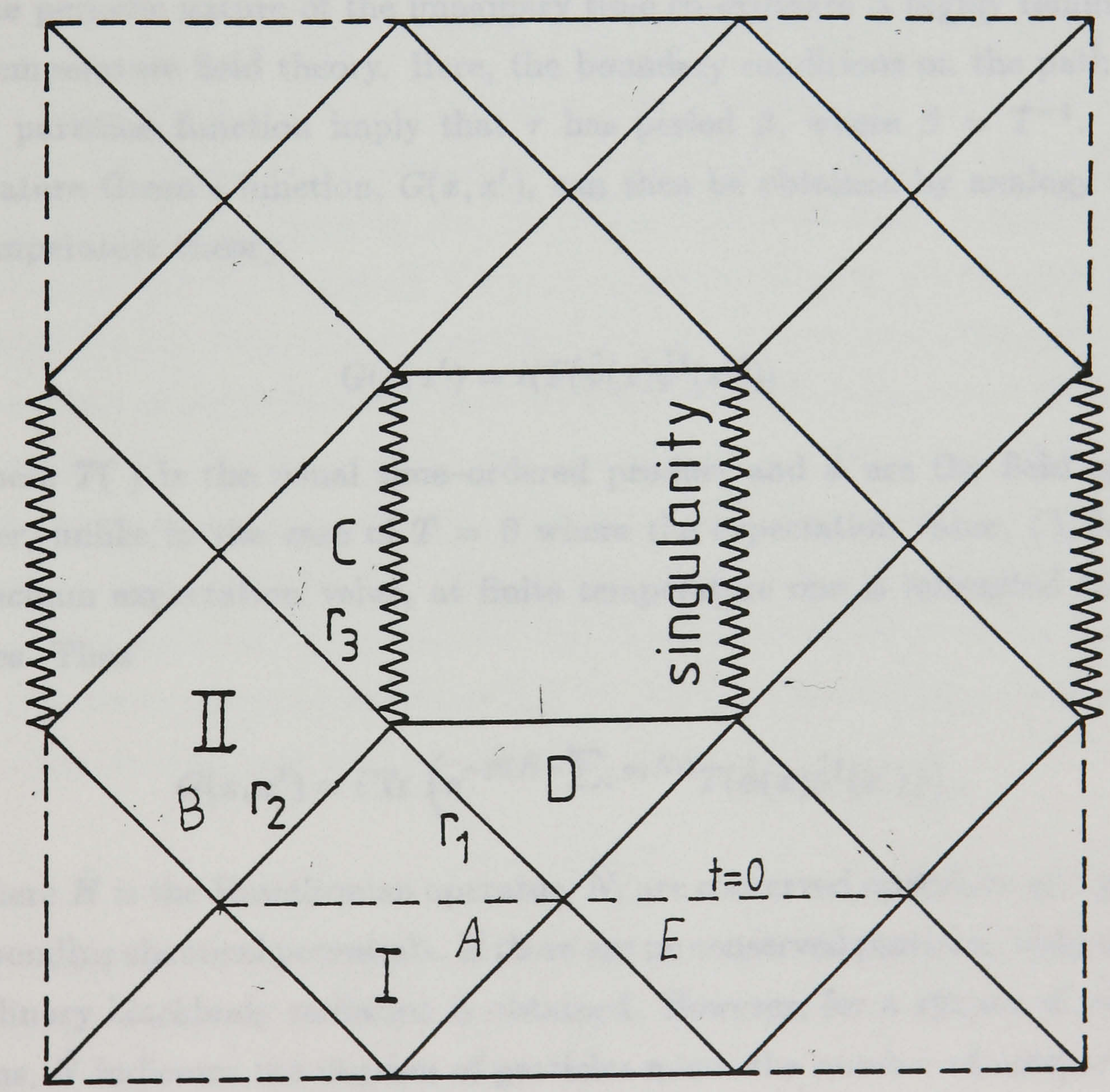

Fig.2.8 Penrose diagram showing time symmetric initial data surface at $t = 0$.

Since a great circle on the Euclidean section goes from r_1, through r_2, back to the same point r_1, the identification with the Lorentzian spacetime implies that the fully extended Penrose diagram must be periodically identified in the spatial direction. This is consistent with the earlier assumption that the spacetime consists of a black hole with two horizons at antipodal points of a closed universe. The matching of the Euclidean section at $\tau = 0, \pi$ with the Lorentzian spacetime at $t = 0$ means that the instanton can be used to calculate the nucleation probability at this initial data surface of a closed universe containing antipodal black hole horizons.

2.4 Chemical Potentials

The periodic nature of the imaginary time co-ordinate is highly reminiscent of finite temperature field theory. Here, the boundary conditions on the path integral for the partition function imply that τ has period β, where $\beta = T^{-1}$. A finite temperature Green's function, $G(x, x')$, can then be obtained by analogy with the zero temperature theory.

$$G(x, x') = i\langle T(\hat{\phi}(x)\hat{\phi}^{\dagger}(x'))\rangle \, , \tag{2.4.1}$$

where $T(\;)$ is the usual time-ordered product and $\hat{\phi}$ are the field operators. However, unlike in the case of $T = 0$ where the expectation value, $\langle\;\rangle$, is defined as a vacuum expectation value, at finite temperature one is interested in thermal averages. Thus

$$G(x, x') = i \operatorname{Tr}\left(e^{-\beta(\hat{H} - \sum_i \mu_i \hat{N}_i)} T(\hat{\phi}(x)\hat{\phi}^{\dagger}(x'))\right) , \tag{2.4.2}$$

where $\hat{H}$ is the Hamiltonian operator, $\hat{N}_i$ are conserved operators and μ_i are the corresponding chemical potentials. If there are no conserved particles, then $\mu = 0$ and the ordinary blackbody radiation is obtained. However, for a system of relativistic fermions, $\hat{N}$ indicates the number of particles minus the number of antiparticles.

The significance of the chemical potential can be seen from the canonical ensemble of statistical mechanics. Just as the energy expectation value is given by differentiating the partition function Z with respect to β,

$$\langle E \rangle = \frac{1}{Z}\frac{dZ}{d\beta} \tag{2.4.3}$$

$$Z = \sum_{\mathrm{states}} e^{-\beta(E-\mu N)} , \tag{2.4.4}$$

so the expectation value of N is given by differentiation with respect to μ,

$$\langle N \rangle = \frac{1}{Z}\frac{dZ}{d\mu} . \tag{2.4.5}$$

The quantity β is equal to the inverse temperature, indicating that bodies at different values of β exchange energy until equilibrium is established and they have the same temperature. A similar interpretation may be given to the chemical potential: bodies at different μ exchange particles to equalise their chemical potentials.

The black hole spacetimes which have been presented here carry two conserved quantities - the charge Q and the angular momentum a. The possibility of thermal radiation through the Hawking process means that both these quantities may be radiated away from the hole itself. This is because the holes have a chemical potential with respect to the surrounding spacetime. It has been seen that chemical potentials appear in the real phase factors of finite temperature Green's functions. Similar phase factors appear for charged particles (e.g. electrons) travelling around the horizon of a charged black hole. This is like the Aharanov–Bohm effect where a charged particle travelling around a solenoid picks up a phase shift equal to $e \oint \underline{A}.\underline{ds}$, even though the $\underline{B}$ field on the trajectory $\underline{ds}$ is zero. For the black hole in de Sitter space with metric (2.2.7) and electromagnetic potential (2.2.10), the potential acting on a charged, but not rotating particle, is in the time direction.

$$A = \frac{iQr}{\rho^2\chi^2} d\tau . \tag{2.4.6}$$

Thus in Euclidean space, an electron travelling around the singularity in the black hole picks up a phase shift which may be identified with a chemical potential. Evaluated at the horizon r_2 this is found to be

$$\mu = \frac{e\,Q\,r_2}{\rho_2^2} = \frac{e\,Q\,(\alpha\chi - \Delta_-)}{(\alpha^2 + a^2 cos^2\theta - \alpha\chi\Delta_- - 2\alpha M\chi^{-1})} . \tag{2.4.7}$$

It should be noted that not only is the evaluation of the chemical potential distance dependent, but there is some ambiguity over whether the singularity in the gauge field which produces the effect is at the black hole event horizon r_2 or the de Sitter horizon r_1. Here, the metric is assumed to have been made regular at r_1, leaving the singularity at r_2.

The resultant effect of this chemical potential is that the black hole will preferentially radiate particles of the same sign as itself in order to achieve a chemical potential equal to that of the surrounding spacetime. Thus the hole discharges.

A similar effect is seen in the case of rotating particles travelling around the horizon of a black hole. Although there is no overall chemical potential if one looks across the whole spectrum of angular momentum, the effect does appear if one picks out a particular mode. The field $\Psi(t, R, \theta, \hat{\phi})$ can be written as

$$\Psi = \sum_n \hat{\Psi}_n e^{in\hat{\phi}} = \sum_n \hat{\Psi}_n e^{in(\phi - \Omega_H t)} \tag{2.4.8}$$

using the transformation of the azimuthal co–ordinate (2.3.2). In Euclidean space with $\tau = it$, the factor $e^{-in\Omega_n t}$ is not periodic, but, as with the charged case, has the form of a phase shift which can be identified with a chemical potential. Thus the rotational chemical potential of a charged, rotating black hole in de Sitter space is

$$\mu = n\Omega_H = n\frac{a}{(r_2^2 + a^2)} \tag{2.4.9}$$

when evaluated at the black hole horizon. Thus the black hole loses angular momentum by preferentially radiating particles rotating in the same direction as itself.

2.5 Euclidean Instantons

It has been shown that a regular Euclidean section can only be obtained when the surface gravity is the same on both horizons. As was seen in §2.2, this occurs in two situations: when the horizons are coincident; and when condition (2.2.21) is met. In the cases with coincident horizons a metric can be defined by taking the limit $r_1 \to r_2$. We then have

$$R^2 = e^{2\kappa r^*} \to \frac{r - r_2}{r_1 - r} \tag{2.5.1}$$

$$\Delta = \alpha^{-2}\frac{R^2}{(1+R^2)^2}(r_2-r_1)^2(r-r_3)(r-r_4)\,. \tag{2.5.2}$$

So the metric has the form

$$\begin{aligned} ds^2 = \frac{4\rho_1^2\alpha^2\chi^2}{(r_1-r_3)(r_1-r_4)}\frac{dR^2+R^2d\Theta^2}{(1+R^2)^2} + \rho_1^2\Delta_\theta^{-1}d\theta^2 \\ + \rho_1^{-2}\Delta_\theta\sin^2\theta\hat{\omega}_1^2\,, \end{aligned} \tag{2.5.3}$$

where r_1 is the repeated root of $\Delta = 0$, $\rho_1 = \rho(r_1)$ and

$$\chi^2\hat{\omega}_1 = -2\hat{a}r_1\frac{(r_1-r_2)}{(r_1^2-\hat{a}^2)}\kappa^{-1}d\Theta - (r_1^2-\hat{a}^2)d\hat{\phi}\,. \tag{2.5.4}$$

Putting $A = \hat{a}/r_1$, $B = \hat{a}/\alpha$ and $C = \hat{Q}/r_1$ gives

$$(r_1-r_3)(r_1-r_4) = r_1^2\frac{A^2}{B^2}\left((1+B^2)+2(A^2+C^2)\right)\,, \tag{2.5.5}$$

and the metric becomes

$$\begin{aligned} ds^2 = 4r_1^2\frac{(1-A^2cos^2\theta)}{(1+B^2)+2(A^2+C^2)}d\Omega_2^2 + r_1^2\frac{(1-A^2cos^2\theta)}{(1-B^2cos^2\theta)}d\theta^2 \\ + r_1^{-2}\frac{(1-B^2cos^2\theta)}{(1-A^2cos^2\theta)}\sin^2\theta\hat{\omega}_1^2\,, \end{aligned} \tag{2.5.6}$$

with

$$\frac{\hat{\omega}_1}{r_1^2} = -\frac{4A}{(1+B^2)+2(A^2+C^2)}\frac{R^2}{(1+R^2)}d\Theta - \frac{(1-A^2)}{(1-B^2)}d\hat{\phi}\,. \tag{2.5.7}$$

In the case of zero charge $B^2 = A^2(1+A^2)/(3-A^2)$. Then the metric becomes

$$\begin{aligned} ds^2 = r_1^2\Bigg\{4\frac{(3-A^2)(1-A^2\cos^2\theta)}{(3+6A^2-A^4)}d\Omega_2^2 + \frac{(3-A^2)(1-A^2\cos^2\theta)}{(3-A^2\sin^2\theta-A^4\cos^2\theta)}d\theta^2 \\ + r_1^{-4}\frac{(3-A^2\sin^2\theta-A^4\cos^2\theta)}{(3-A^2)(1-A^2\cos^2\theta)}\sin^2\theta\hat{\omega}_1^2\Bigg\}\,, \end{aligned} \tag{2.5.8}$$

with

$$\frac{\hat{\omega}_1}{r_1^2} = -\frac{4A(3-A^2)}{(3+6A^2-A^4)}\frac{R^2}{(1+R^2)}d\Theta - \frac{(1-A^2)(3-A^2)}{(3-2A^2-A^4)}d\hat{\phi}\,. \tag{2.5.9}$$

This is equivalent to the Page $S_2 \times S_2$ metric [56].

The only remaining case with equal surface gravities are specified by the condition

$$M^2 = \chi^2 P^2 - \chi^2 \hat{Q}^2 - \chi^4 \hat{a}^2 . \tag{2.5.10}$$

The mass must be positive for a real metric, and therefore the magnetic charge P has to be non–zero.

The contributions of these instantons to quantum transitions are proprtional to e^{-I}, where I is the instanton action. The present system has an Einstein–Maxwell action

$$I = -\frac{1}{16\pi} \int \left(R - 2\Lambda - F_{ab} F^{ab} \right) \sqrt{g}\, d^4x . \tag{2.5.11}$$

Substituting the metric and vector potential for the instantons corresponding to rotating magnetically charged black holes with separate horizons leads to

$$I = -\pi\alpha^2 + 2\pi\alpha M^3/\chi^3 \left(M^2 - \hat{a}^2 \chi^4 \right) . \tag{2.5.12}$$

The instanton action for black holes which are not rotating but which carry both electric and magnetic charge is

$$I = -2\pi\alpha^2 + \pi\alpha \, (-M^2 - P^2 + \hat{Q}^2)/M , \tag{2.5.13}$$

with

$$P^2 - \hat{Q}^2 = M^2 . \tag{2.5.14}$$

In the case of $\hat{Q} = P = \hat{a} = 0$, both these examples reduce to a four sphere of radius α with action $I_0 = -\pi\alpha^2$. In the same limit, the instanton corresponding to black holes in de Sitter space with co–incident horizons r_1 and r_2 has an action of only $\frac{2}{3} I_0$. The coincident horizon instantons are therefore suppressed in comparison to those with separate horizons.

2.6 Applications

If $\hat{Q} = \hat{a} = 0$, the instantons provide the amplitudes for the pair creation of magnetically charged black holes. If the amplitudes are small, these pairs are nucleated at a rate per unit volume of

$$N = A e^{-B} . \tag{2.6.1}$$

The rate is normalised with respect to the state with no black holes present, where the action is that of a four–sphere. Thus $B = I - I_0$. Since the cosmological constant has dimensions of L^{-2}, the factor A may be estimated to be α^4. Hence

$$N \sim \alpha^{-4} e^{-2\pi\alpha M} \ . \tag{2.6.2}$$

Such black holes could possibly appear as relic monopoles from an inflationary era in which the universe undergoes a de Sitter phase of exponential expansion. The density of such monopoles at the end of the de Sitter phase is, by (2.6.2),

$$\rho_M \sim M\alpha^{-3} e^{-2\pi\alpha M} \ . \tag{2.6.3}$$

This can be compared with the density of photons

$$\rho_\gamma = \frac{3}{8\pi}\alpha^{-2} \ . \tag{2.6.4}$$

Since in a radiation dominated universe ρ_γ varies with $a^4(t)$, where $a(t)$ is the scale factor, and matter scales as $a^3(t)$, the ratio of monopoles to photons should now be

$$\left.\frac{\rho_M}{\rho_\gamma}\right|_{\mathrm{now}} = (1+z)^{-1} \left.\frac{\rho_M}{\rho_\gamma}\right|_{\mathrm{deSitter}} \sim (1+z)^{-1} \frac{8\pi}{3} M\alpha^{-1} e^{-2\pi\alpha M} \ , \tag{2.6.5}$$

where z is the redshift related to the scale factor by $(1+z)^{-1} = a(t_{\mathrm{now}})/a(t_{\mathrm{deS}})$. The redshift can be expressed in terms of the temperatures of the radiation at the end of the de Sitter phase and now:

$$(1+z)^{-1} = \frac{T_\gamma(t_{\mathrm{deS}})}{T_\gamma(t_{\mathrm{now}})} = \frac{\alpha^{-1/2}}{10^{-32}} \ . \tag{2.6.6}$$

Thus

$$\left.\frac{\rho_M}{\rho_\gamma}\right|_{\mathrm{now}} \sim 10^{33} \alpha^{-3/2} M e^{-2\pi\alpha M} \ . \tag{2.6.7}$$

Or, expressed as a fraction of the critical density, the monopole abundance is

$$\Omega_M \sim 10^{29} \alpha^{-3/2} M e^{-2\pi\alpha M} \ . \tag{2.6.8}$$

This rules out examples with $M < \alpha/4 \leq 1.65$, since from (2.2.22) $M < \alpha/4$ for real radii and from observation the total $\Omega \leq 1$. Such small values of α would normally lead to large density fluctuations, but this has only been demonstrated for inflationary models driven by scalar fields [57].

The instantons discussed above should also contribute to spacetime foam calculations in which quantum wormholes of the order of the Planck scale connect the universe to other 'large' universes. An example of such a calculation is Coleman's explanation of the vanishing cosmological constant [19]. Here Coleman employs the notion that spacetime may consist of many universes joined via wormholes to derive a theory in which the cosmological constant plays the role of a dynamical variable. This theory yields a probability distribution which is infinitely peaked on the submanifold with zero cosmological constant. The argument may be followed in principle using the approach of Klebanov, Banks and Susskind [58]. If two points x, x' are connected by a wormhole, then the generating functional defining a low energy field theory is given by

$$Z = \int \delta\phi\, \delta g \, \exp\left\{ -I(g, \lambda) + \tfrac{1}{2} C_{ij} \int d^4x \sqrt{g}\, \theta_i(x) \int d^4x' \sqrt{g'}\, \theta_j(x') \right\} , \qquad (2.6.9)$$

where λ indicates all parameters such as couplings, masses and the cosmological constant Λ. $\theta_i(x)$ are the local operators at x and C_{ij} is a coupling matrix. Thus the first term in the exponent is the action at one point of the manifold, whilst the second term represents the effective interaction due to the presence of a wormhole. Noting the identity

$$\begin{aligned} &\exp\left\{ -\tfrac{1}{2} C^{-1}_{ij} \left(\alpha_i + C_{ik} \int d^4x \sqrt{g}\, \theta_k(x) \right) \left(\alpha_j + C_{jl} \int d^4x' \sqrt{g'}\, \theta_l(x') \right) \right\} \\ &= \exp\left\{ -\tfrac{1}{2} C_{ij} \alpha_i \alpha_j - \alpha_i \int d^4x \sqrt{g}\, \theta_i(x) - \tfrac{1}{2} C_{ij} \int d^4x \sqrt{g}\, \theta_i(x) \int d^4x' \sqrt{g'}\, \theta_j(x') \right\} \end{aligned} \qquad (2.6.10)$$

where the left hand side is simply a Gaussian, (2.6.9) can be rewritten

$$Z = \int \delta\phi\, \delta g \, e^{-(\lambda_i + \alpha_i) \int d^4x \sqrt{g} \theta_i(x)} . \int d\alpha \, e^{-1/2 C^{-1}_{ij} \alpha_i \alpha_j} . \qquad (2.6.11)$$

Thus the variables α_i appear as coupling constants.

However, (2.6.11) only describes the situation in which a wormhole joins two distant points of the same universe, like a handle. One must also take into account wormholes which join two or more universes (fig.2.9).

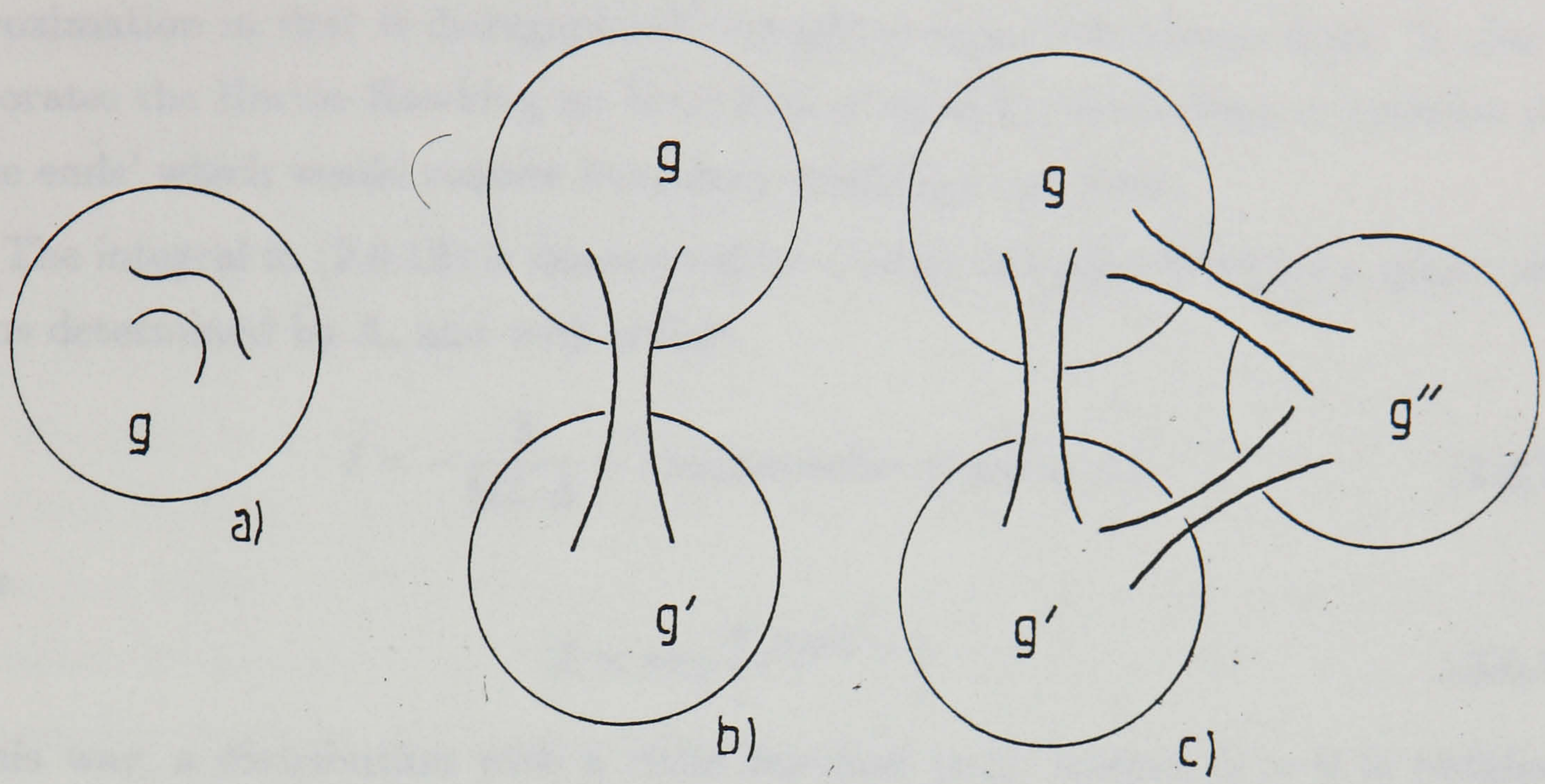

Fig.2.9 Some wormholes.

In this case the generating functional again takes the form

$$Z(\alpha_i) = \int \delta\phi\, \delta g\, e^{-(\lambda_i+\alpha_i)\int d^4x\sqrt{g}\theta_i(x)} . \int d\alpha\, \rho(\alpha)\,, \tag{2.6.12}$$

but with

$$\rho(\alpha) = e^{-1/2 C_{ij}^{-1}\alpha_i\alpha_j} \exp\left\{\int \delta g' e^{-(\lambda_k+\alpha_k)\int d^4x\sqrt{g}\theta_k(x)}\right\} . \tag{2.6.13}$$

The form of this probability distribution can be justified by expanding the second exponential:

$$\rho(\alpha) = e^{-1/2\, C_{ij}^{-1}\alpha_i\alpha_j}\left\{1 + \int \delta g' e^{-(\lambda_k+\alpha_k)\int d^4x\sqrt{g}\,\theta_k(x)} + \frac{1}{2}\int \delta g' \int \delta g'' e^{-(\lambda_k+\alpha_k)\int d^4x\sqrt{g}\,\theta_k(x)} e^{-(\lambda_l+\alpha_l)\int d^4x\sqrt{g}\,\theta_l(x)} + \ldots\right\} . \tag{2.6.14}$$

The first term of (2.6.14) describes the interaction of a handle (fig.2.9a), the second term that of a wormhole joining two universes (fig.2.9b), the third that of a wormhole joining three universes (fig.2.9c), etc.. The calculation assumes the dilute gas approximation in that it disregards all wormhole–wormhole interactions. It also incorporates the Hartle–Hawking no boundary propsal by disallowing wormholes with 'loose ends' which would require boundary conditions on them.

The integral in (2.6.12) is dominated by a single universe given by a sphere with radius determined by Λ, and with action

$$I = -\frac{3}{8G\Lambda} + (\text{higher order terms in } \Lambda) \ . \tag{2.6.15}$$

Thus

$$Z \propto \exp\left\{e^{3/8G\Lambda}\right\} \ . \tag{2.6.16}$$

In this way, a distribution with a delta function peak around $\Lambda = 0$ is obtained. However, (2.6.15) assumes an expansion in integral powers of Λ. This contrasts with the instantons discussed above which have actions with terms of order $\Lambda^{-1/2}$ ((2.5.12) and (2.5.13)). Although inclusion of terms with non–integral powers of Λ in the integration over all instantons will alter the form of the effective action, it will have no effect on the vanishing of the cosmological constant since this depends only on the leading order term. However, the double exponential of (2.6.13) implies that the probability distribution is also peaked around the minima of the higher order terms in (2.6.15). This is the 'big fix' which fixes all the constants of the theory. Since these higher order terms depend on the nature of the expansion employed, the big fix will be affected by the presence of terms of non–integral powers in Λ.

It seems strange that black holes with no magnetic charge P, but with separate horizons, do not correspond to any gravitational instantons. The complexification of the electric charge, Q, and angular momentum, a, in the condition (2.5.10) for equal surface gravities forces the rest mass, M, to become imaginary. This is physically unacceptable. The alternative, if one is to succeed in finding an instanton corresponding to uncharged holes, is not to complexify the rotation parameter and to learn to live with complex metrics. Such a complex metric will nevertheless have the following properties:

(i) the manifold is compact;

(ii) the metric coefficients are smooth functions;

(iii) the Einstein equations are satisfied;

(iv) the determinant of the metric is real and positive;

(v) the action is real.

These properties seem to be consistent with a reasonable choice of saddle points and they could be used to define a generalised gravitational instanton. The first two points are conditions for the metric to lie on a contour of integration of the path integral. The third condition means that the metric is a saddle point and the remaining conditions imply an interpretation of the amplitude in terms of a tunnelling event. As was mentioned in §2.1 similar considerations with regard to complex contours have arisen in the context of quantum cosmology. In the light of this work, it is perhaps possible to interpret the instantons which have been described in this chapter as providing the probability of transition from the Euclidean regime to a classical Lorentzian spacetime containing the appropriate black holes in de Sitter space. The maximal slice of the Euclidean sections would provide the initial conditions of that classical spacetime. Moreover, this interpretation is valid for instantons derived from complex metrics, so that the unattractive necessity of non–zero magnetic charge is avoided.

CHAPTER THREE

STABILITY OF BLACK HOLES IN DE SITTER SPACE

3.1 Perturbation Calculations

The issue of the stability of black holes was raised in the previous chapter. There it was seen that the global structure of metrics like the Reissner–Nordström metric raised the possibility of an observer utilising the wormhole properties of the black hole and following a timelike curve from one asymptotic region of the universe to another. This would involve crossing the black hole horizons. At the outer black hole event horizon this need not present a problem. However, the inner horizon is a Cauchy horizon representing the limit of predictability from some initial data. Travel beyond such a horizon is problematic – causality is violated and naked singularities are encountered. Physics completely fails to tell us how to deal with such conditions. These conceptual difficulties are avoided if the Cauchy horizon is unstable so that it cannot actually be traversed. Then the unpleasant features of the region to the future of the horizon are forever hidden from the view of any observer originating in the region to the past of the horizon.

The stability of the Cauchy horizon can only be examined by a full gravitational perturbation analysis. Early perturbation calculations [59] began with the aim of understanding the final state of gravitational collapse. It was known that the ideal case of the collapse of a spherically symmetric body in the vacuum would result in a Schwarzschild black hole. What was not known was whether the collapse of a body which deviated slightly from sphericity would evolve towards the Schwarzschild metric. The perturbed Einstein equations derived from a Schwarzschild background metric with first order perturbations can be separated into modes using spherical harmonics. If the frequencies of these modes are real, then the metric is stable; if imaginary then it is unstable. It is found that there are no modes of imaginary frequency with acceptable behaviour at infinity. Thus there is no exponential behaviour and the spacetime is stable to first order deviations from spherical symmetry.

The separability of the perturbation equations implies that they reduce to problems of scattering. This picture was developed by later calculations [21,60]. The theory of potential scattering describes waves incident upon some potential using solutions to a Schrödinger equation. For instance, for a one–dimensional potential, $V(x)$, solutions are sought to the equation

$$\left(\frac{d^2}{dx^2} + \sigma^2\right) f = Vf \; , \quad (-\infty < x < \infty) \, . \tag{3.1.1}$$

The asymptotic behaviour of solutions to (3.1.1), for $x \to \pm\infty$, is given by $e^{\pm i\sigma x}$. Through examining a wave with this asymptotic form incident upon an appropriate potential and giving rise to a reflected wave and a transmitted wave, it is possible to discover whether a black hole is stable in the sense that perturbations remain bounded as they evolve.

As mentioned above, the stability of the Reissner–Nordström metric is of particular interest due to the global structure of the spacetime. As in the Schwarzschild case, an observer could fall through the outer horizon by following a future–directed timelike trajectory and become forever lost from the sight of an external onserver. At the instant of crossing the horizon, any radiation emitted by the observer would appear infinitely redshifted to the external obsever. Although the traveller appears to take an infinite time to fall into the hole as viewed from outside, their own experience is of a continued journey within the black hole. Here the differences between Reissner–Nordström and Schwarzschild black holes become apparant. In the latter case any observer who had crossed the horizon at $r = 2M$ would inevitably fall towards the singularity at $r = 0$. For a charged black hole, this fate is avoidable since timelike paths exist which do not terminate on the singularity. However, at the instant of crossing the Cauchy horizon, the observer would see radiation which had a finite frequency outside the hole, as having an infinite frequency. In other words, there is an infinite blue–shift effect which results in the observer witnessing the entire future of the universe in only a finite time interval. This suggests that the Cauchy horizon is unstable against perturbations in the initial data which give rise to waves, and that such perturbations would in general lead to singularities on the horizon. Thus the qualitative features of the Reissner–Nordström spacetime do, after all, appear to be similar to those of the Schwarzschild: an obsever crossing the outer horizon inevitably proceeds towards a physical singularity.

The analysis of the stability of the Reissner–Nordström metric was made rigorous by Chandrasekhar and Hartle [20] using the methods of potential scattering. The radiation generated by gravitational perturbations outside the black hole either cross the event horizon directly or after having been scattered by the curvature of the background spacetime. The existence of poles in the reciprocal transmission coefficient of this scattered radiation results in a flux which is divergent on the Cauchy horizon. The same behaviour is also found in the case of a Kerr geometry since this displays the same features as the Reissner–Nordström metric: namely, a timelike singularity and a Cauchy horizon. In this case, however, the perturbation analysis is somewhat more lengthy – a note in reference21at the end of detailed calculations on perturbations of the Kerr metric explains that going from one step to the next can entail as many as fifty extra pages of computation. In the present work, attention is restricted to Reissner–Nordström black holes!

The question to be addressed in the following is how the stability of a black hole is effected if the hole is situated in de Sitter space. As was seen in the previous chapter, the presence of a cosmological constant alters the structure of the spacetime in such a way that the future null infinity, $\Im^+$, of flat space, is replaced by a de Sitter event horizon. Thus an observer at the Cauchy horizon, rather than seeing a flash of radiation of infinite energy density, sees only the finite number of events which lie within the cosmological horizon. It will be shown in this chapter that for a black hole in de Sitter space, a zero in the amplitude of the perturbations entering the hole cancel with the relevant pole in the transmission coefficient. The cosmological constant therefore renders the Cauchy horizon stable.

3.2. No Hair Theorems

In addition to these perturbative results there are a number of results that are known as 'no hair' theorems. Broadly, these state that all stationary black holes are axially symmetric and are characterised by their mass, charge and angular momentum alone. Together with the results of the perturbation calculations, these theorems give a picture of realistic gravitational collapse leading to the formation of black holes with the Kerr–Newman metric. This assumes the validity of the cosmic censorship conjecture which will be brought into question by the results of §3.5.

The no hair theorems are a series of results derived by a number of authors between 1967 and 1975. They essentially fall into four steps associated with a number of other proofs which relax the assumptions of the original theorems, or which generalise the vacuum results to the electrovac case. These steps are as follows:

1. Hawking [61]. Each connected component of a two–surface representing the intersection of the past event horizon with the future event horizon of a stationary black hole has the topology of a two–sphere. This is demonstrated by studying the convergent properties of future directed outgoing null geodesics orthogonal to the surface when that surface is deformed outwards into the region exterior to the hole. Thus it is expected that any stationary black hole have a spherical boundary.
2. Hawking [61]. All stationary black holes must be either static or axisymmetric. This is shown by consideration of the Killing vectors of the spacetime. Since the time translation isometry leaves the horizon of a stationary black hole invariant, the Killing vector field $\partial/\partial t$ must be tangent to the horizon and therefore always be either spacelike or null there. However, there is a one parameter transformation which leaves the Cauchy data of the horizons invariant but which moves points on the horizons along the null generators of the horizon. Thus there must exist a Killing vector field which, on the horizons, is directed along the null generators. If no ergosphere is present, the stationary Killing field is timelike or null everywhere outside the black hole. Thus it must be null on the horizon, so the spacetime is static. If an ergosphere is present, then $\partial/\partial t$ becomes spacelike outside the hole and is therefore spacelike on the horizon. The required Killing vector which is null on the horizon must then be distinct from $\partial/\partial t$. By taking a linear combination of these two vector fields it is possible to consruct a Killing vector reflecting a second isometry of the spacetime. Since the orbits of this vector field are closed, it represents an axial Killing field. Thus the spacetime is axisymmetric.
3. Israel [62]. The only black holes which are static and topologically spherical are the Schwarzschild solutions (or the Reissner–Nordström solutions in the electrovac generalisation).

4. Carter [63]. All stationary axisymmetric black holes are uniquely characterised by only two parameters (three for electrovac spacetimes). Since the Kerr solutions exhaust all possible values of these parameters, these must be the only stationary axisymmetric black holes.

The theorems of Hawking show that all stationary solutions must be spherical and either static or axisymmetric. The other theorems demonstrate that the known solutions are the only ones satisfying these conditions. Thus the no hair theorems illustrate the simplicity of black holes. Taken together with the idea of cosmic censorship, they imply that the gravitational collapse of a sufficiently massive body will always result in a black hole with the metric outside the collapsing body described by one of the Kerr–Newman family of solution; and furthermore, that the only physical properties of the body of relevance to the final state of collapse are the mass, charge and angular momentum. The results of perturbative analyses provide the mechanism whereby this reduction to only three parameters may be understood. All other initial characteristics of the collapsing body are carried away by gravitational radiation produced on the surface of an otherwise spherical body.

As was discussed in chapter two, de Sitter spacetime can be thought of as sharing some of the properties of black hole metrics due to the presence of an event horizon. One can, for instance, assign a surface gravity to the de Sitter horizon and thereby examine the thermodynamics of the spacetime, just as one can for black holes. It is natural, therefore, to ask whether there is a theorem for de Sitter space which is in some way analogous to the no hair theorems of black holes. Roughly speaking such a 'theorem' would state that all spacetimes with a positive cosmological constant would approach the de Sitter solution asymptotically in time. The cosmic no hair conjecture was first proposed by Gibbons and Hawking [53], but has been difficult to prove or even formalise. Indeed, there are obvious counter examples, such as the closed Friedmann–Robertson–Walker universe which recollapses rather than relaxes to de Sitter. However, the concept has been pursued, in suitably weakened statements, due to its importance for the inflationary scenario.

Inflation assumes that the universe has a phase during which it is trapped in a false vacuum whose energy dominates the evolution of the universe at that time. The vacuum energy plays the role of a repulsive cosmological constant, Λ. If the geometry of spacetime then approaches that of de Sitter space, the universe will undergo exponential expansion at a Hubble rate $H = \sqrt{\Lambda/3}$. If that period of exponential expansion is of sufficient length, i.e. about sixty Hubble times, then the homogeneity and isotropy of the observable universe can be explained. Inflation serves to decrease the dependence of a reasonable cosmological model on the initial conditions. It is no longer necessary to assume Friedman–Robertson–Walker symmetries from the outset. It is not obvious, however, that models with alternative initial conditions will actually enter the de Sitter epoch required by inflation. It is for this reason that some sort of cosmic no hair principle is desired. This would ensure the naturalness of inflation in freeing models from any specific assumptions of homogeneity or isotropy.

A proof of the cosmic no hair conjecture has been given [64] for the Bianchi models which represent all homogeneous, but anisotropic, cosmologies. The Hamiltonian constraint equation and Raychaudhuri equation for such models enable one to place limits on the shear of a timelike normal and the energy momentum tensor of matter, in terms of the volume expansion rate. These limits indicate that in one expansion time, $\sqrt{3/\Lambda}$, the spacetime approaches an isotropic, homogeneous vacuum spacetime with a volume expansion rate $K = \sqrt{3\Lambda}$. i.e. de Sitter space is achieved. Attempts have been made to confirm the cosmic no hair conjecture for inhomogeneous spacetimes. These have been less successful and are summarised in reference 65. Some inhomogeneous models support the conjecture whilst others do not.

The mechanism whereby cosmic baldness is realised can be investigated via perturbation techniques. Thus the stability of de Sitter spacetimes, both with and without black holes, is of interest; the former for their relevance in the inflationary context, the latter for the implications of travel across the Cauchy horizon. Before either of these situations can be examined, the perturbation equations governing the response of a general black hole de Sitter spacetime to gravitational and electromagnetic perturbations must be set up.

3.3 The Perturbation Equations

The analysis of the stability of a de Sitter black hole to gravitational perturbations will follow closely the procedure of Chandrasekhar [21] in his study of the Reissner–Nordström solution. In accordance with his notation, the metric of a charged black hole in de Sitter space (2.2.7) can be rewritten in the form

$$ds^2 = e^{2\nu}dt^2 - e^{2\varphi}d\phi^2 - e^{2\mu_2}dr^2 - e^{2\mu_3}d\theta^2 \ , \tag{3.3.1}$$

where

$$e^{2\nu} = \frac{\Delta}{r^2} \ ; \quad e^{2\mu_2} = \frac{r^2}{\Delta} \ ; \quad e^{2\mu_3} = r^2 \ ; \quad e^{2\varphi} = r^2 \sin^2\theta \tag{3.3.2}$$

and

$$\Delta = -\tfrac{1}{3}\Lambda r^4 + r^2 - 2Mr + Q^2 \ . \tag{3.3.3}$$

Since the metric is spherically symmetric, only the axisymmetric modes of the perturbations need be considered. Modes which lack axial symmetry and have a ϕ dependence $e^{im\phi}$ can be deduced from suitable rotations of the axisymmetric $m = 0$ modes. These modes decompose under rotation into non–axisymmetric components which represent the different non–axisymmetric modes. For this reason attention is restricted to perturbations with (t, r, θ) dependence only.

The perturbed metric is taken to be

$$ds^2 = e^{2\nu}dt^2 - e^{2\varphi}(d\phi - w dt - q_2 dr - q_3 d\theta)^2 - e^{2\mu_2}dr^2 - e^{2\mu_3}d\theta^2 \ . \tag{3.3.4}$$

The perturbations are given by the quantities w, q_2, q_3 and the small increments $\delta\nu$, $\delta\mu_2$, $\delta\mu_3$, $\delta\varphi$ of the functions ν, μ_2, μ_3, φ respectively. These perturbations fall into two orthogonal classes, as can be seen by observing the effect of a change of the sign of ϕ. A reversal in the sign of ϕ has no effect on the set $\delta\nu$, $\delta\mu_2$, $\delta\mu_3$ and $\delta\varphi$; so these represent polar perturbations. The quantities w, q_2 and q_3, on the other hand, must also undergo a change in sign if the metric is to remain uneffected; thus these represent axial perturbations. The two types of perturbation decouple in that they can be treated independently of each other.

The relevant perturbation equations are obtained by linearising the field equations about the solution (3.3.1). Since the black hole is charged, this involves the linearisation of the Maxwell equations as well as the Einstein equations. The only non–vanishing component of the unperturbed Maxwell tensor is

$$F_{tr} = -Q/r^2 \ . \tag{3.3.5}$$

All other components are of first order in perturbations.

For convenience, a tetrad formalism is adopted rather than the usual local co-ordinate system. An orthonormal basis of four linearly independent vector fields with signature (+,–,–,–) is chosen and relevant quantities are projected onto this basis. The covariant vectors of the tetrad frame are given by

$$\begin{aligned} e_{(0)\,i} &= (e^{\nu}, 0, 0, 0) \ , \\ e_{(1)\,i} &= (we^{\varphi}, -e^{\varphi}, q_2\, e^{\varphi}, q_3\, e^{\varphi}) \ , \\ e_{(2)\,i} &= (0, 0, -e^{\mu_2}, 0) \ , \\ e_{(3)\,i} &= (0, 0, 0, -e^{\mu_3}) \ . \end{aligned} \tag{3.3.6}$$

The linearised Maxwell equations fall into two sets of three equations: those resulting from the axial perturbations,

$$\begin{aligned} (re^{\nu} F_{01} \sin\theta)\,_{,r} + re^{-\nu} F_{12,0} \sin\theta &= 0 \ , \\ re^{\nu} \left(F_{01} \sin\theta\right)_{,\theta} + r^2 F_{13,0} \sin\theta &= 0 \\ re^{-\nu} F_{01,0} + \left(re^{\nu} F_{12}\right)_{,r} + F_{13,\theta} &= -\sin\theta\, Q\, \left(w_{32} - q_{2,0}\right) \ ; \end{aligned} \tag{3.3.7}$$

and those resulting from the polar perturbations ,

$$\begin{aligned} -\left(re^{\nu} F_{23}\right)_{,r} + re^{-\nu} F_{03,0} &= 0 \ , \\ \left(re^{\nu} F_{23} \sin\theta\right)_{,\theta} + \left(e^{\varphi+\mu_3} F_{02}\right)_{,0} &= 0 \ , \\ \left(e^{\mu_2+\nu} F_{02}\right)_{,\theta} - \left(re^{\nu} F_{03}\right)_{,r} + re^{-\nu} F_{23,0} &= 0 \ . \end{aligned} \tag{3.3.8}$$

The perturbed Einstein equations are given by

$$\delta R_{ab} = -2[\eta^{nm}(\delta F_{an} F_{bm} + F_{an} \delta F_{bm}) - \eta_{ab} Q \delta F_{tr}/r^2] \ . \tag{3.3.9}$$

which again separate into axial and polar equations.

For both axial and polar perturbations the time dependence varies as $e^{i\sigma t}$.

The axial perturbations of the Einstein equations give

$$\left(A\,(r,\theta)\right)_{,\theta} + r^4 \left(w_{,r} - q_{2,t}\right)_{,t} \sin^3\theta = 4Qre^{\nu} B\,(r,\theta) \tag{3.3.10}$$

$$\left(A\,(r,\theta)\right)_{,r} + r^2 e^{-2\nu} \left(w_{,\theta} - q_{3,t}\right)_{,t} \sin^3\theta = 0 \ , \tag{3.3.11}$$

and the first three Maxwell equations (3.3.7) give

$$\left(e^{2\nu}\left(re^{\nu}B\left(r,\theta\right)\right)_{,r}\right)_{,r}+\frac{e^{\nu}}{r}\left(\frac{B\left(r,\theta\right)_{,\theta}}{\sin\theta}\right)_{,\theta}\sin\theta-re^{-\nu}B\left(r,\theta\right)_{,t,t} \tag{3.3.12}$$
$$=Q\left(w_{,r,t}-q_{2,t,t}\right)\sin^{2}\theta\ ,$$

where

$$A(r,\theta)=\Delta(q_{2,\theta}-q_{3,r})\sin^{3}\theta \tag{3.3.13}$$

and

$$B(r,\theta)=F_{t\phi}\sin\theta\ . \tag{3.3.14}$$

It is possible to further separate variables by introducing

$$A(r,\theta)=A(r)\,C_{l+2}^{-3/2}(\theta)\ , \tag{3.3.15}$$

$$B(r,\theta)=3B(r)\,C_{l+1}^{-1/2}(\theta)\ , \tag{3.3.16}$$

where $C_{l+2}^{-3/2}$ is the Gegenbauer function related to the Legendre function $P_l(\theta)$ by the formula

$$C_{l+2}^{-3/2}(\theta)=(P_{l,\theta,\theta}-P_{l,\theta}\cot\theta)\sin^{2}\theta \tag{3.3.17}$$

and satisfying the equation

$$\frac{1}{\sin\theta}\frac{dC_{\lambda}^{\nu}}{d\theta}=-2\nu C_{\lambda-1}^{\nu+1}\ . \tag{3.3.18}$$

Then equations (3.3.10) to (3.3.12) reduce to

$$r^{2}e^{2\nu}\left(\frac{e^{2\nu}A_{,r}}{r^{2}}\right)_{,r}-\mu^{2}\frac{e^{2\nu}}{r^{2}}A+\sigma^{2}A=-\frac{4Q}{r}\mu^{2}e^{3\nu}B \tag{3.3.19}$$

and

$$\left(e^{2\nu}\left(re^{2\nu}B\right)_{,r}\right)_{,r}-\left(\mu^{2}+2\right)\frac{e^{\nu}}{r}B+\left(\sigma^{2}re^{-\nu}-\frac{4Q^{2}}{r^{3}}e^{\nu}\right)B=-\frac{Q}{r^{4}}A\ , \tag{3.3.20}$$

where

$$\mu^{2}=2n=(l-1)(l+2)\ . \tag{3.3.21}$$

Equations (3.3.19) and (3.3.20) may be decoupled using the substitutions

$$
\begin{aligned}
Z_1^- &= p_1 H_1^- + (-p_1 p_2)^{\frac{1}{2}} H_2^- \\
Z_2^- &= -(-p_1 p_2)^{\frac{1}{2}} H_1^- + p_1 H_2^- \ ,
\end{aligned}
\tag{3.3.22}
$$

where

$$
\begin{aligned}
H_1^- &= -2\mu r e^{\nu} B \\
H_2^- &= \frac{A(r)}{r} \ ,
\end{aligned}
\tag{3.3.23}
$$

and

$$
\begin{aligned}
p_1 &= 3M + (9M^2 + 4Q^2\mu^2)^{\frac{1}{2}} \\
p_2 &= 3M - (9M^2 + 4Q^2\mu^2)^{\frac{1}{2}} \ .
\end{aligned}
\tag{3.3.24}
$$

This yields two 1-dimensional Schrödinger-type wave equations of the form

$$
D^2 Z_i^- = V_i^- Z_i^- \ , \quad (i = 1, 2) \ , \tag{3.3.25}
$$

with potentials given by

$$
V_i^- = \frac{\Delta}{r^5} \left[(\mu^2 + 2)\, r - p_j \left(1 + \frac{p_i}{\mu^2 r} \right) \right] \ , \quad (i, j = 1, 2; i \neq j) \ , \tag{3.3.26}
$$

and differential operator defined by

$$
D^2 = \frac{d^2}{dr^{*2}} + \sigma^2 \ . \tag{3.3.27}
$$

Here the tortoise co-ordinate r^* has been introduced and is defined as in chapter two, i.e.

$$
\frac{dr^*}{dr} = \frac{r^2}{\Delta} \ . \tag{3.3.28}
$$

A similar series of steps may be followed for the polar perturbations, although the algebra involved is this time more lengthy and so the substitutions required are more complicated. The variables r, θ can be separated by the substitutions

$$\begin{aligned}
\delta\nu &= N(r)P_l \\
\delta\mu_2 &= L(r)P_l \\
\delta\mu_3 &= T(r)P_l + V(r)P_{l,\theta,\theta} \\
\delta\varphi &= T(r)P_l + V(r)P_{l,\theta}\cot\theta \\
\delta F_{tr} &= \frac{r^2 e^{2\nu}}{2Q} B_{tr}(r)P_l \\
F_{t\theta} &= -\frac{re^{\nu}}{2Q} B_{t\theta}(r)P_{l,\theta} \\
F_{r\theta} &= -\frac{i\sigma re^{-\nu}}{2Q} B_{r\theta}(r)P_{l,\theta} \ ,
\end{aligned} \tag{3.3.29}$$

where $P_l(\theta)$ are the Legendre polynomials. This yields the following set of equations for the radial functions:

$$\begin{aligned}
N_{,r} &= aN + bL + c(nV - B_{r\theta}) \\
L_{,r} &= (a - \frac{1}{r} + \nu_{,r})N + (b - \frac{1}{r} - \nu_{,r})L + c(nV - B_{r\theta}) - \frac{2}{r}B_{r\theta} \\
nV_{,r} &= -(a - \frac{1}{r} + \nu_{,r})N - (b + \frac{1}{r} - 2\nu_{,r})L - (c + \frac{1}{r} - \nu_{,r})(nV - B_{r\theta}) + B_{t\theta} \\
B_{r\theta,r} &= B_{t\theta} - \frac{2}{r}B_{r\theta} \\
B_{t\theta,r} &= 2e^{-2\nu}Q^2\frac{(N+L)}{r^3} - 2\left(\frac{1}{r} + \nu_{,r}\right)B_{t\theta} - B_{tr} - \sigma^2 e^{-4\nu}B_{r\theta} \ , \\
B_{tr} &= 2r^{-4}e^{-2\nu}Q^2\left[2T - l(l+1)V\right] - l(l+1)r^{-2}e^{-2\nu}B_{r\theta}
\end{aligned} \tag{3.3.30}$$

where

$$\begin{aligned}
a &= \frac{(n+1)}{r}e^{2\nu} \\
b &= -\frac{1}{r} + \nu_{,r} + r\nu_{,r}^2 + \sigma^2 e^{-4\nu}r - \frac{2e^{-2\nu}}{r^3}Q^2 - \frac{ne^{-2}\nu}{r} \\
c &= -\frac{1}{r} + r\nu_{,r}^2 + \sigma^2 e^{-4\nu}r - \frac{2e^{-2\nu}}{r^3}Q^2 + \frac{e^{-2\nu}}{r} \ .
\end{aligned} \tag{3.3.31}$$

As in the case for the axial perturbations, it is possible to decouple these equations by a suitable choice of substitutions. By analogy with the axial equations put

$$\begin{aligned}
Z_1^+ &= p_1 H_1^+ + (-p_1 p_2)^{\frac{1}{2}} H_2^+ \\
Z_2^+ &= -(p_1 p_2)^{\frac{1}{2}} H_1^+ + p_1 H_2^+ \ ,
\end{aligned} \tag{3.3.32}$$

where p_1 and p_2 are as above (3.3.24), but H_1^+ and H_2^+ are given by

$$\begin{aligned} H_1^+ &= -\frac{1}{Q\mu}\left[r^2 B_{r\theta} + 2Q^2\frac{r}{\varpi}(L + nV - B_{r\theta})\right] \\ H_2^+ &= rV - \frac{r^2}{\varpi}(L + nV - B_{r\theta}) \ , \end{aligned} \tag{3.3.33}$$

where

$$\varpi = nr + 3M - \frac{2Q^2}{r} \ . \tag{3.3.34}$$

After some remarkable cancellations, two 1-D Schrödinger equations are again obtained:

$$D^2 Z_i^+ = V_i^+ Z_i^+ \ , \quad (i = 1, 2) \ , \tag{3.3.35}$$

having potentials

$$\begin{aligned} V_1^+ &= \frac{\Delta}{r^5}\left[U + \tfrac{1}{2}(p_1 - p_2)W\right] \\ V_2^+ &= \frac{\Delta}{r^5}\left[U - \tfrac{1}{2}(p_1 - p_2)W\right] \ , \end{aligned} \tag{3.3.36}$$

where

$$\begin{aligned} U &= (2nr + 3M)W + \left(\varpi - nr - M - \tfrac{2}{3}\Lambda r^3\right) - \frac{2nr^2}{\varpi}e^{2\nu} \\ W &= \frac{re^{2\nu}}{\varpi^2}(2nr + 3M) + \frac{1}{\varpi}\left(nr + M + \tfrac{2}{3}\Lambda r^3\right) \ . \end{aligned} \tag{3.3.37}$$

In the above calculation, a system of five linear equations (3.3.30) with five variables, has been reduced to just two second order equations for particular combinations, Z_i, of the five variables. Such a reduction seems remarkable, especially given the complexity of the potentials derived; but it is always possible if some particular solutions of the equations are known beforehand. One might then ask: can the particular solutions which made possible a reduction of this sort be found if this reduction were carried out in an empirical fashion? There does exist an algorism which will fulfil this task [66], and it has been applied to solutions of the form (3.3.35) in reference 21. It is then possible to construct general expressions for the variables $N, L, V, B_{t\theta}, B_{r\theta}$ in terms of the particular solutions and linear combinations of H_1^+, H_2^+ and their derivatives. From these expressions and the substitutions (3.3.29), the behaviour of

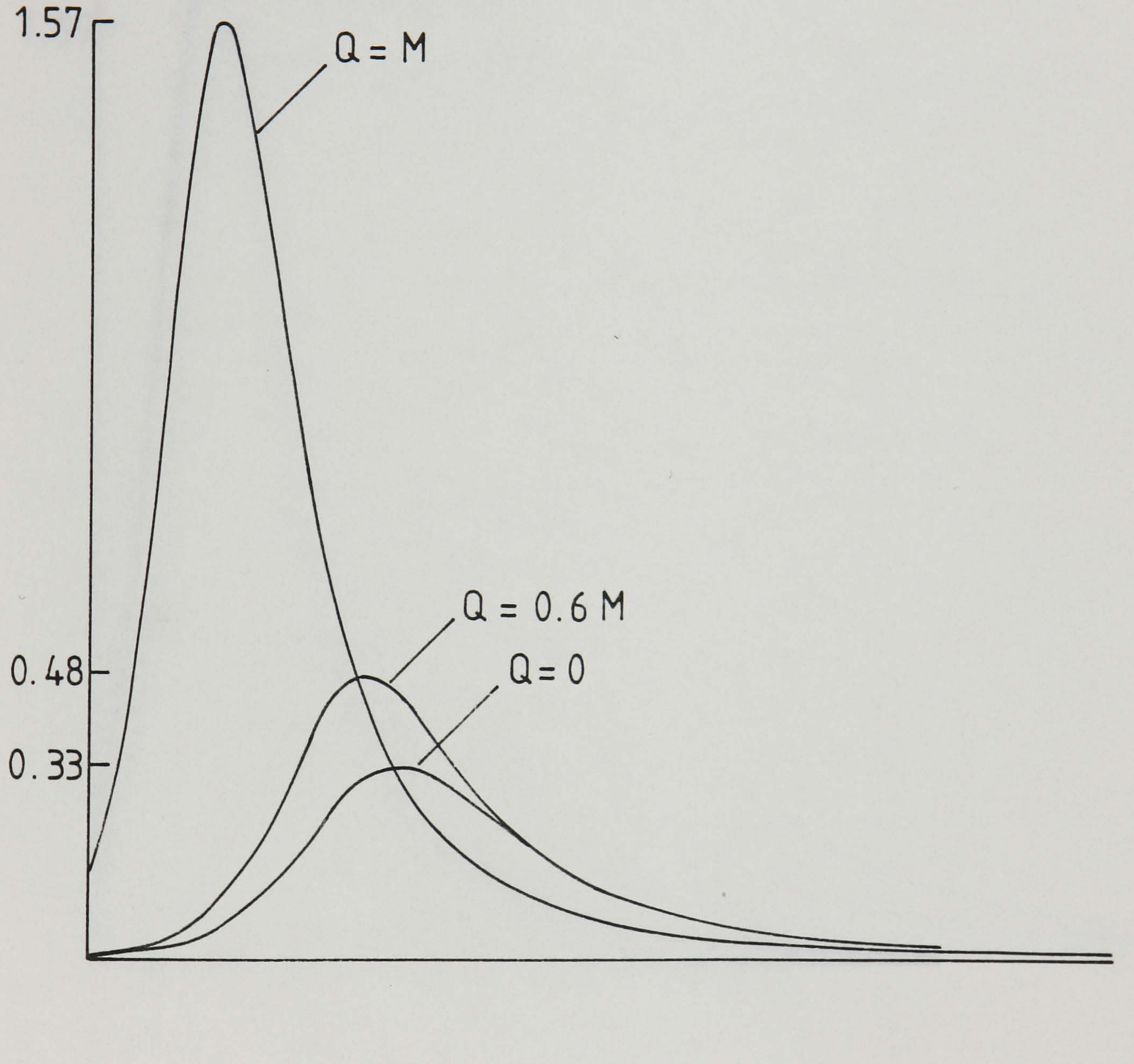

Fig. 3.1 Plot of V_1^+ from $r = 1.44$ to 7.00 with $\alpha = 8M$, $n = 0$

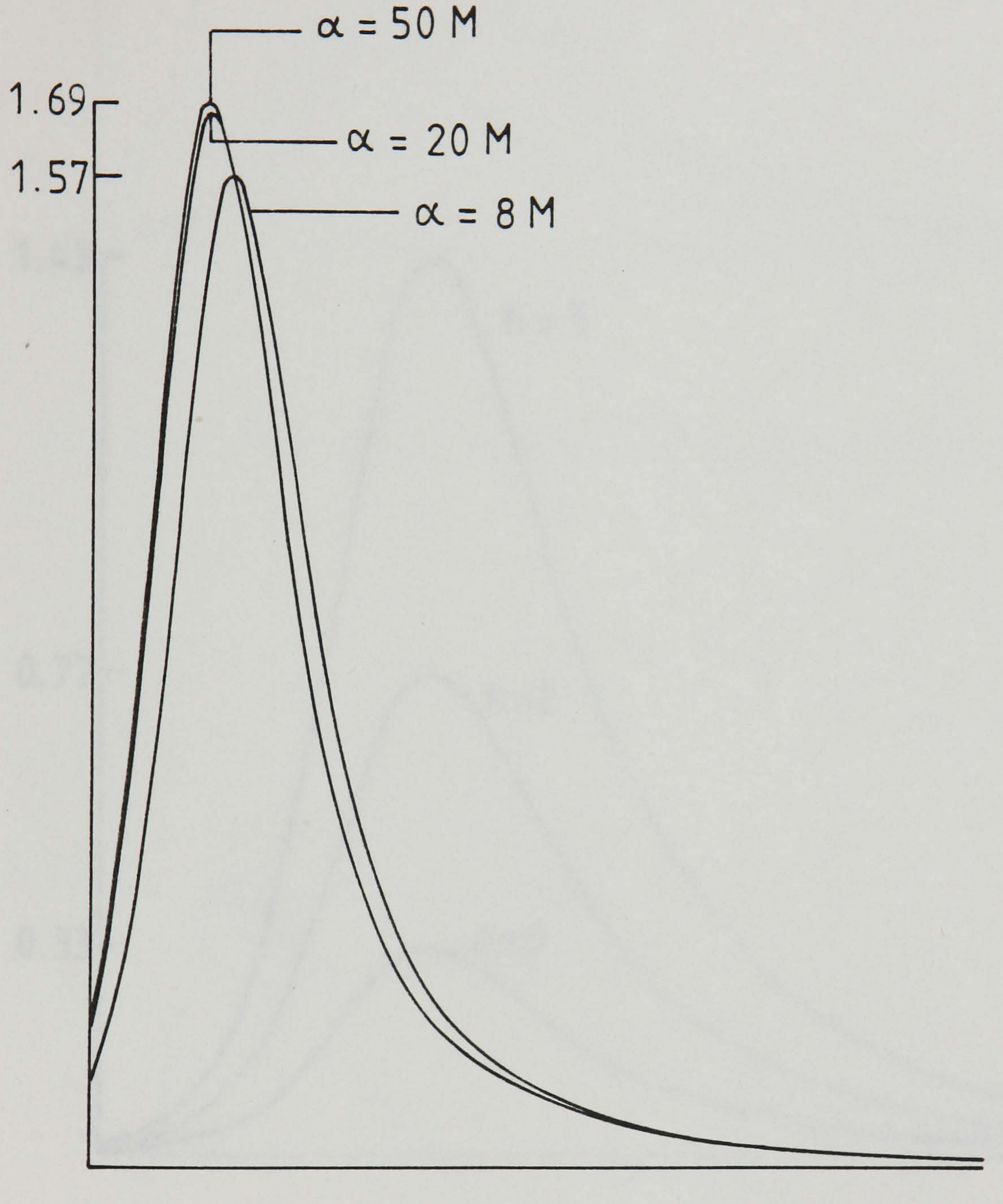

Fig. 3.2 Plot of V_1^+ from $r = 1.02$ to 7.00 with $Q = M$, $n = 0$

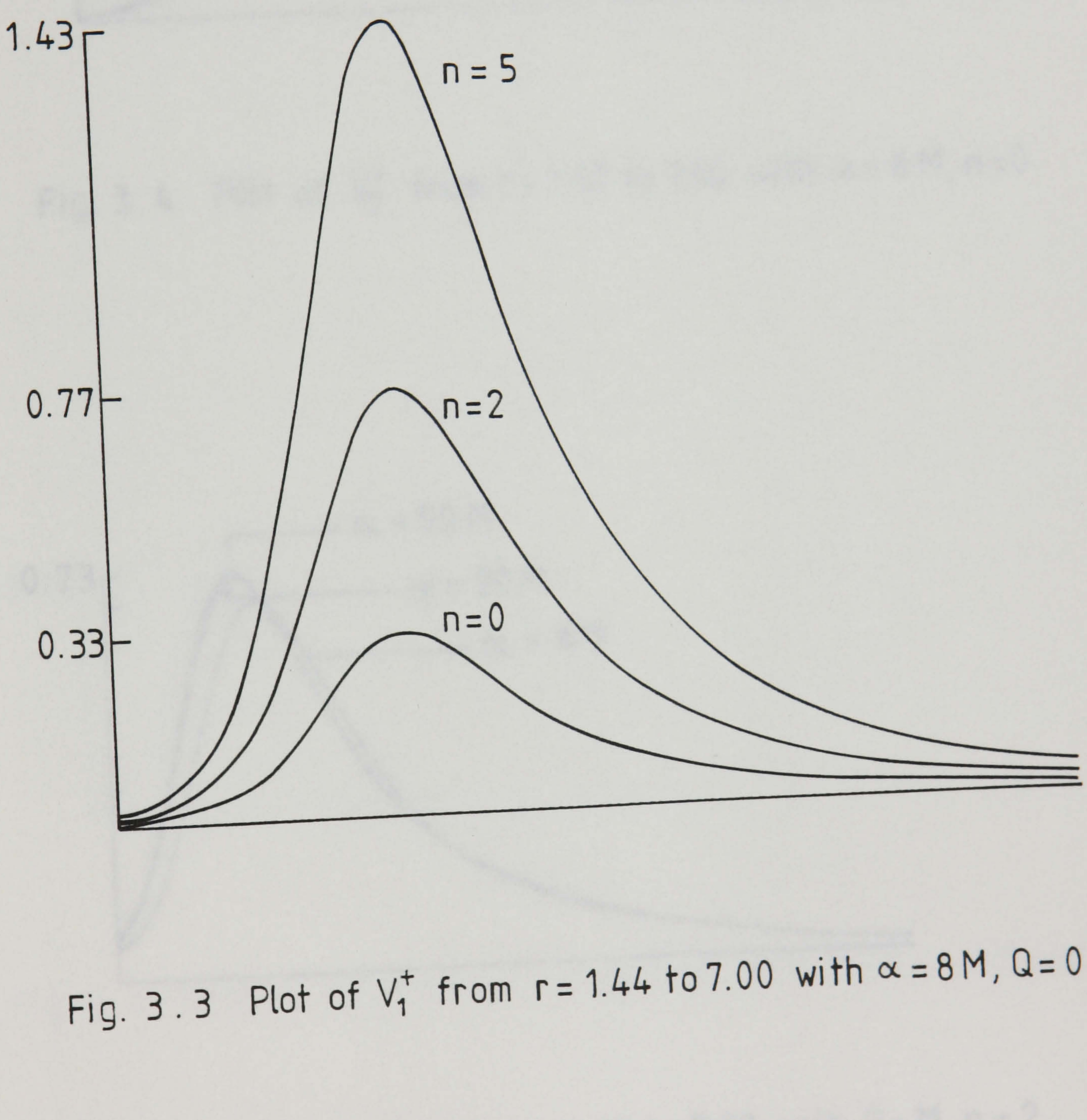

Fig. 3 . 3 Plot of V_1^+ from $r = 1.44$ to 7.00 with $\alpha = 8M$, $Q = 0$

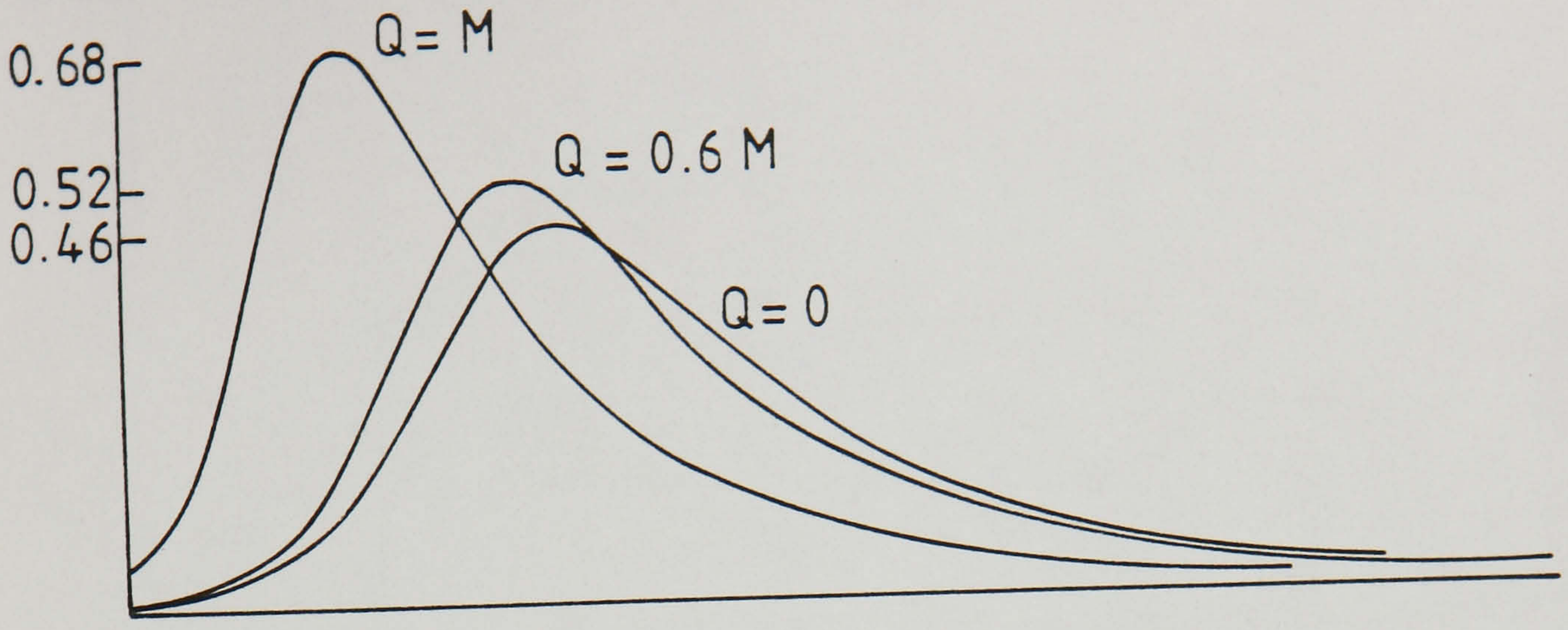

Fig. 3.4 Plot of V_2^+ from r = 1.02 to 7.00 with α = 8 M, n = 0

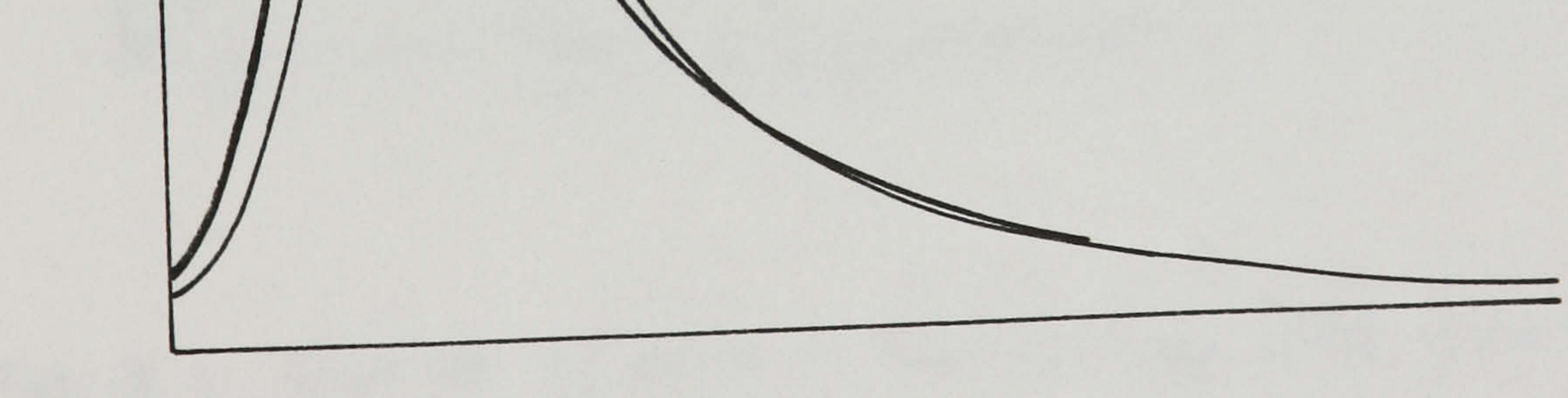

Fig. 3.5 Plot of V_2^+ from r = 1.02 to 7.00 with Q = M, n = 2

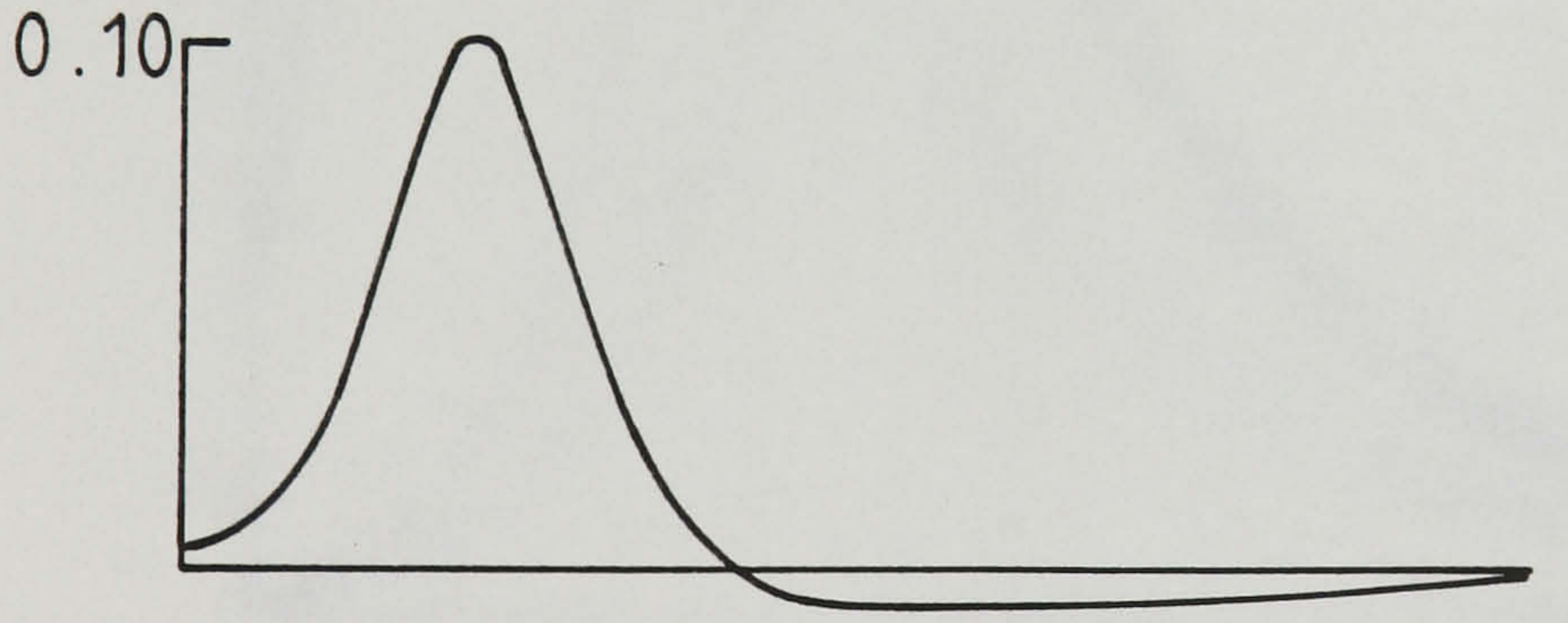

Fig. 3.6 Plot of V_2^+ from $r = 1.44$ to 7.50 with $Q = 0, \alpha = 8M$ and $n = 0$

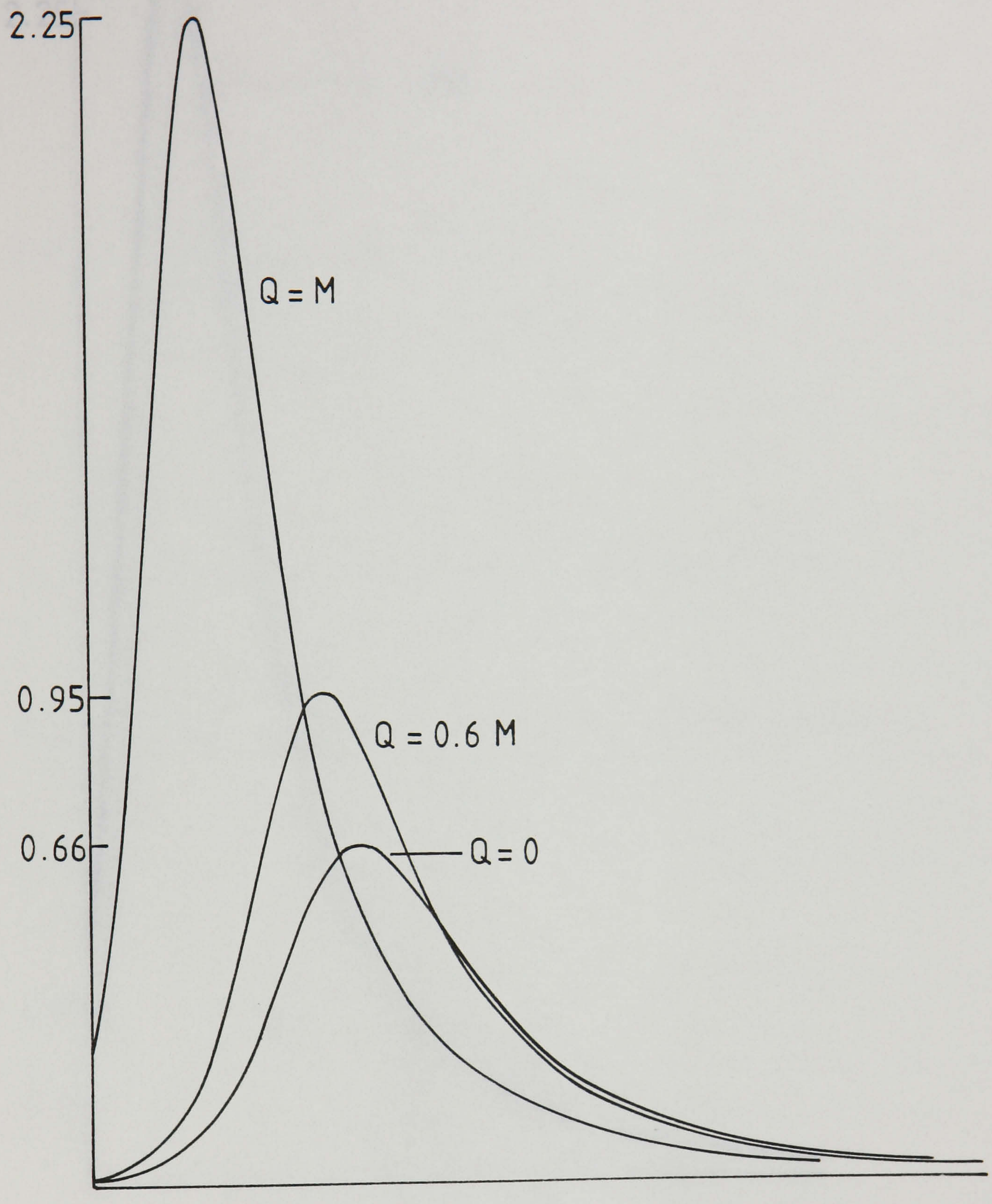

Fig. 3.7 Plot of V_1^- from $r = 1.02$ to 7.00 with $\alpha = 8M$, $n = 2$

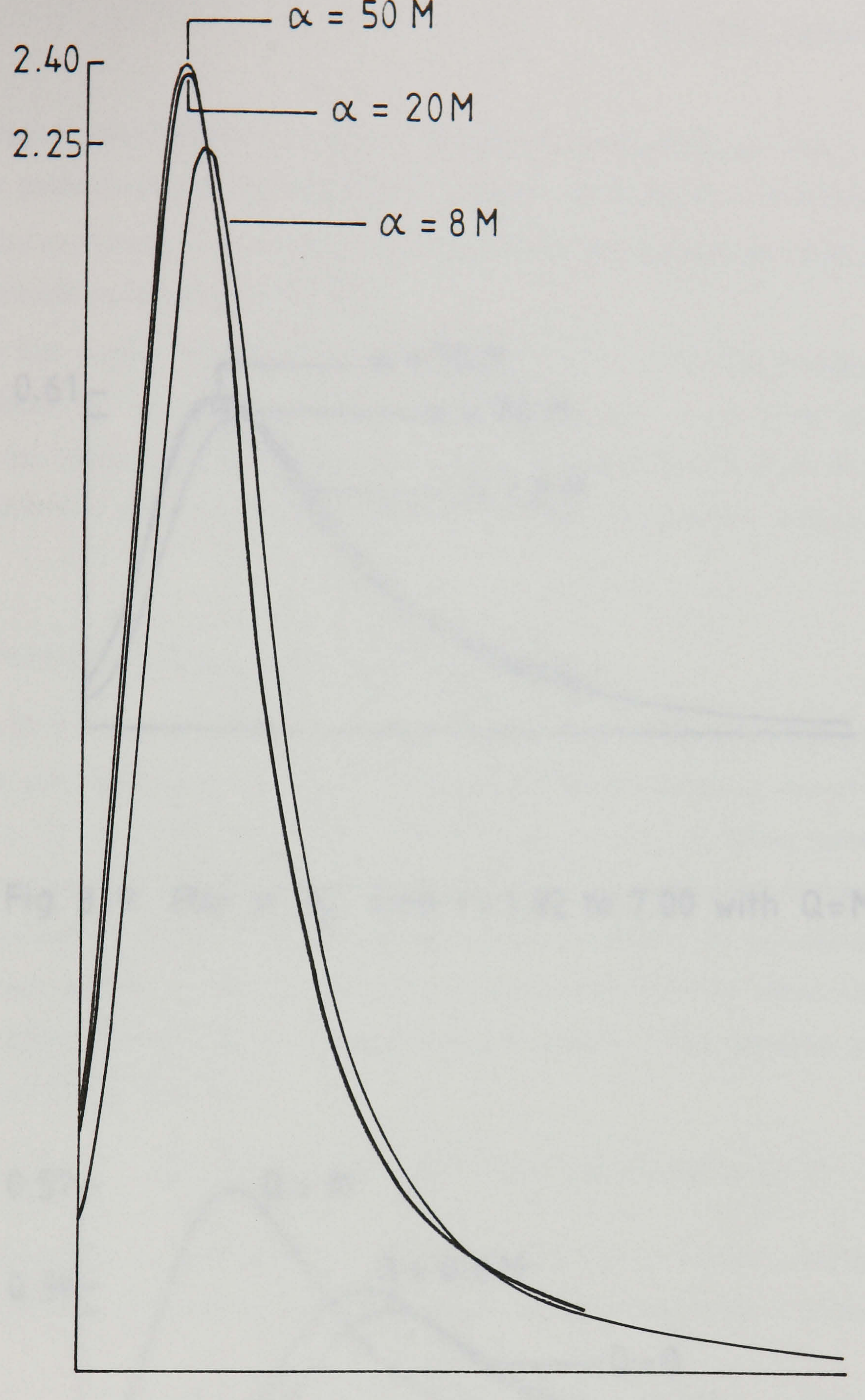

Fig. 3.8 Plot of V_1^- from r = 1.02 to 7.00 with Q = M, n = 2

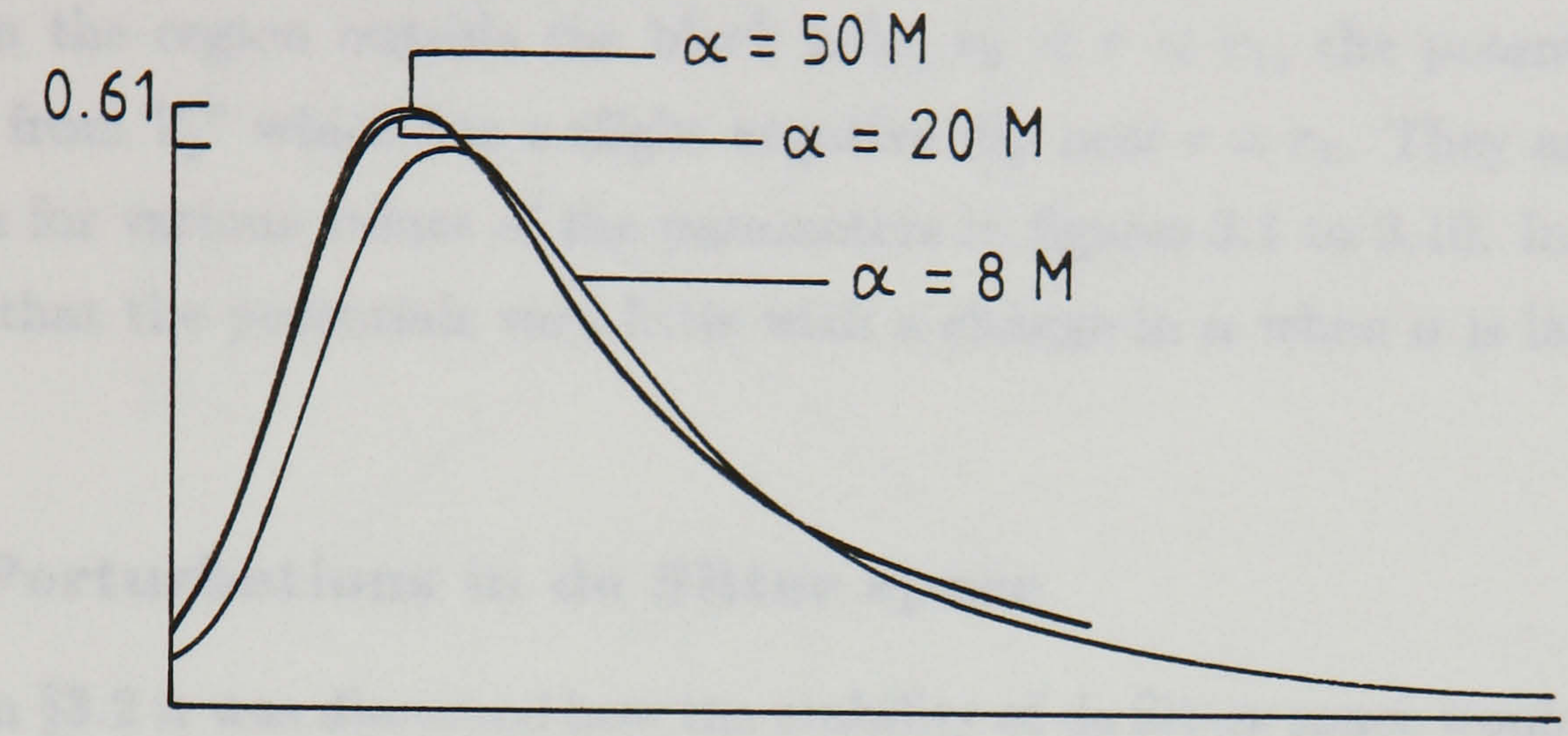

Fig. 3.9 Plot of V_2^- from $r = 1.02$ to 7.00 with $Q = M$, $n = 2$

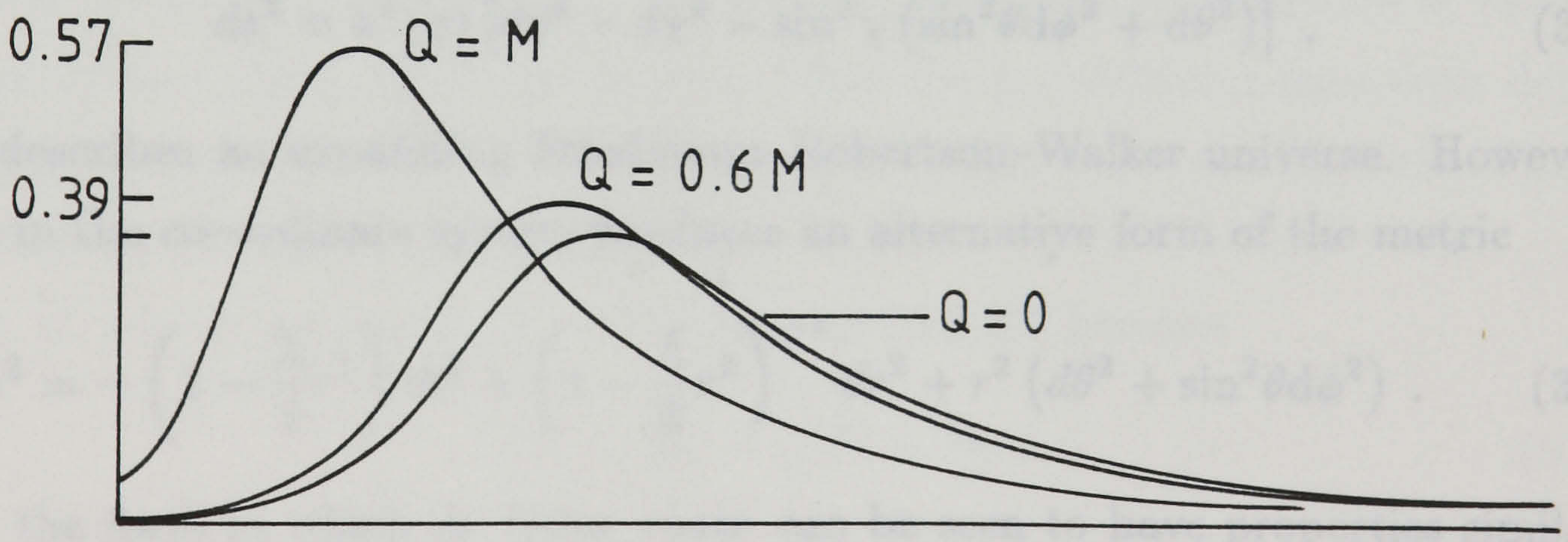

Fig. 3.10 Plot of V_2^- from $r = 1.02$ to 7.00 with $\alpha = 8M$, $n = 2$

the original gravitational and electromagnetic perturbations can be ascertained. A similar procedure may be applied to the case of the axial perturbations.

The evolution of perturbations is therefore completely described by a set of one–dimensional scattering problems.

In the region outside the black hole, $r_2 < r < r_1$, the potentials are positive apart from V_2^+ which has a slight negative dip near $r = r_1$. They are plotted in this region for various values of the parameters in figures 3.1 to 3.10. In particular these show that the potentials vary little with a change in α when α is large.

3.4 Perturbations in de Sitter space

In §3.2 it was discussed how the stability of de Sitter space would reflect a cosmic no hair principle which is important in ensuring the naturalness of the inflationary scenario. It is possible to analyse the perturbations of de Sitter space using conventional cosmological perturbation theory [67]. Small gravitational wave perturbations, decomposed in terms of spherical harmonics, are found to asymptotically approach a constant amplitude as the scale factor approaches infinity. Since the perturbations fail to grow at late times, the spacetime is stable. This analysis expresses the de Sitter metric in the form

$$ds^2 = a^2(\eta)\left[d\eta^2 - d\chi^2 - \sin^2\chi\left(\sin^2\theta \mathrm{d}\phi^2 + \mathrm{d}\theta^2\right)\right] , \tag{3.4.1}$$

which describes an expanding Friedmann–Robertson–Walker universe. However, a change in the co–ordinate system produces an alternative form of the metric

$$ds^2 = -\left(1 - \frac{\Lambda}{3}r^2\right)dt^2 + \left(1 - \frac{\Lambda}{3}r^2\right)^{-1} dr^2 + r^2\left(d\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right) . \tag{3.4.2}$$

This is the form in which de Sitter space can be seen to have properties similar to those of black holes; the universe is static with an event horizon. Figure 3.1 shows the conformal diagram of this spacetime. The diagonal lines are the cosmological event horizons of an observer at $r = 0$. Boucher and Gibbons [68] have shown that the behaviour of perturbations in de Sitter space can be seen most clearly when the metric is expressed in the form (3.4.2). In this frame perturbations are radiated

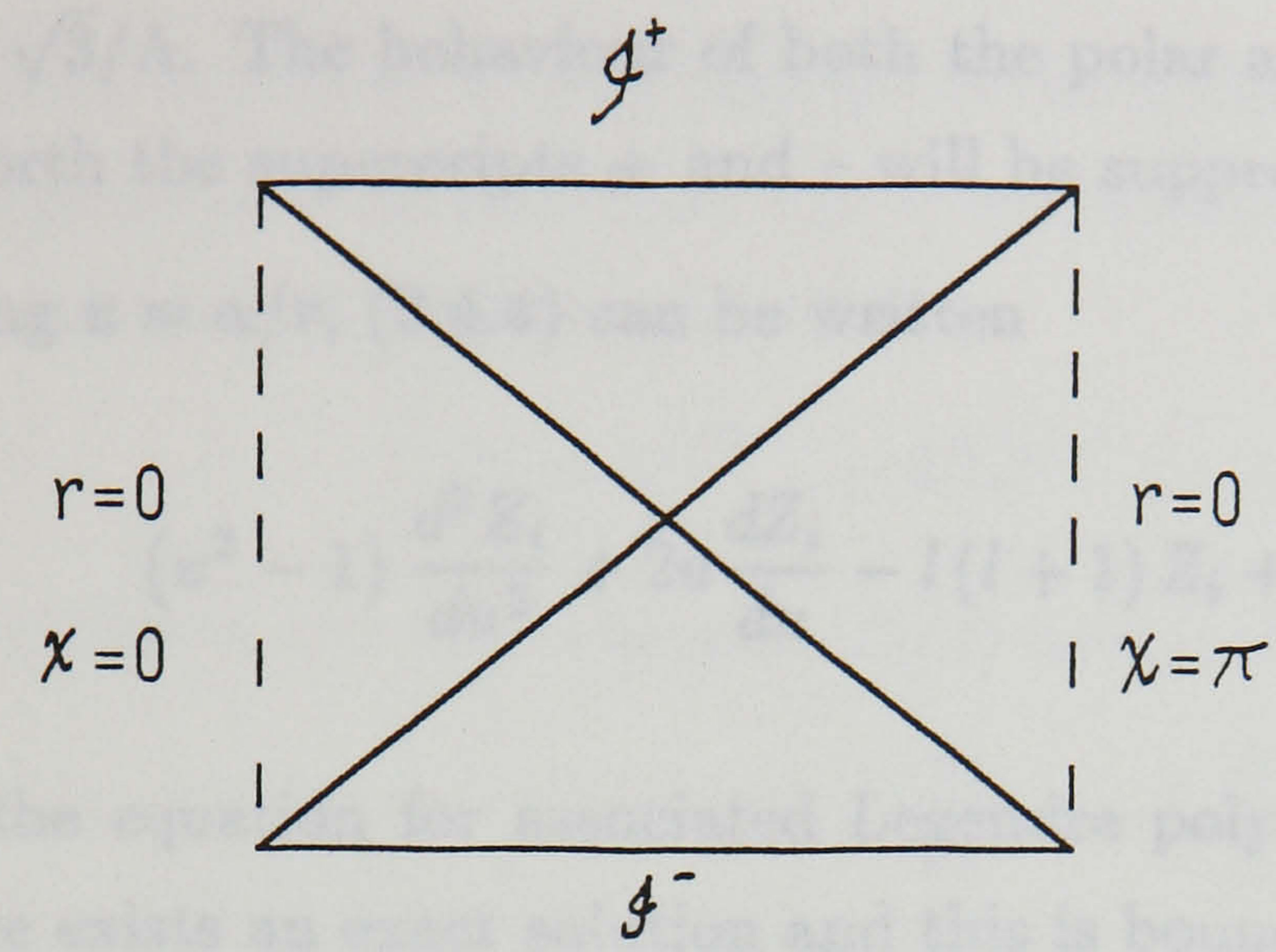

Fig.3.1 Penrose diagram of de Sitter space, showing past and future cosmological event horizons for observer at $r = 0$.

through the cosmological event horizon as they are through the outer horizon of a black hole.

Metric (3.3.1) reduces to empty de Sitter space in the co-ordinates of (3.4.2) if the mass M and charge Q are set to zero. Thus the perturbation equations derived in the previous section from the perturbation of (3.3.1) allow the Boucher-Gibbons approach to be pursued further.

For M and Q tending to zero, (3.3.22) and (3.3.32) become

$$Z_1^{(\pm)} = H_1; \quad Z_2^{(\pm)} = H_2 , \tag{3.4.3}$$

and obey the simple equations

$$\frac{d^2 Z_i^{\pm}}{dr^{*2}} + \sigma^2 Z_i^{\pm} = \frac{l(l+1)\Delta}{r^4} Z_i^{\pm} , \tag{3.4.4}$$

where

$$\Delta = r^2 - r^4/\alpha^2 \tag{3.4.5}$$

and

$$r = \alpha \tanh(r^*/\alpha) , \tag{3.4.6}$$

with $\alpha = \sqrt{3}/\Lambda$. The behaviour of both the polar and the axial modes is the same, so henceforth the supercripts $+$ and $-$ will be suppressed.

Setting $u = \alpha/r$, (3.4.4) can be written

$$(u^2 - 1)\frac{d^2 Z_i}{du^2} + 2u\frac{dZ_i}{du} - l\,(l+1)\,Z_i + \frac{\alpha^2\sigma^2}{(u^2-1)} Z_i = 0 , \tag{3.4.7}$$

which is the equation for associated Legendre polynomials $P_l^m(\alpha/r)$, if $m = i\alpha\sigma$. Thus there exists an exact solution and this is bounded on the horizon.

An observer at $r = 0$ sees waves enter region I of figure 3.2, where $0 \leq r < \alpha$, through the past cosmological horizon, pass through $r = 0$ and leave through the future cosmological horizon.

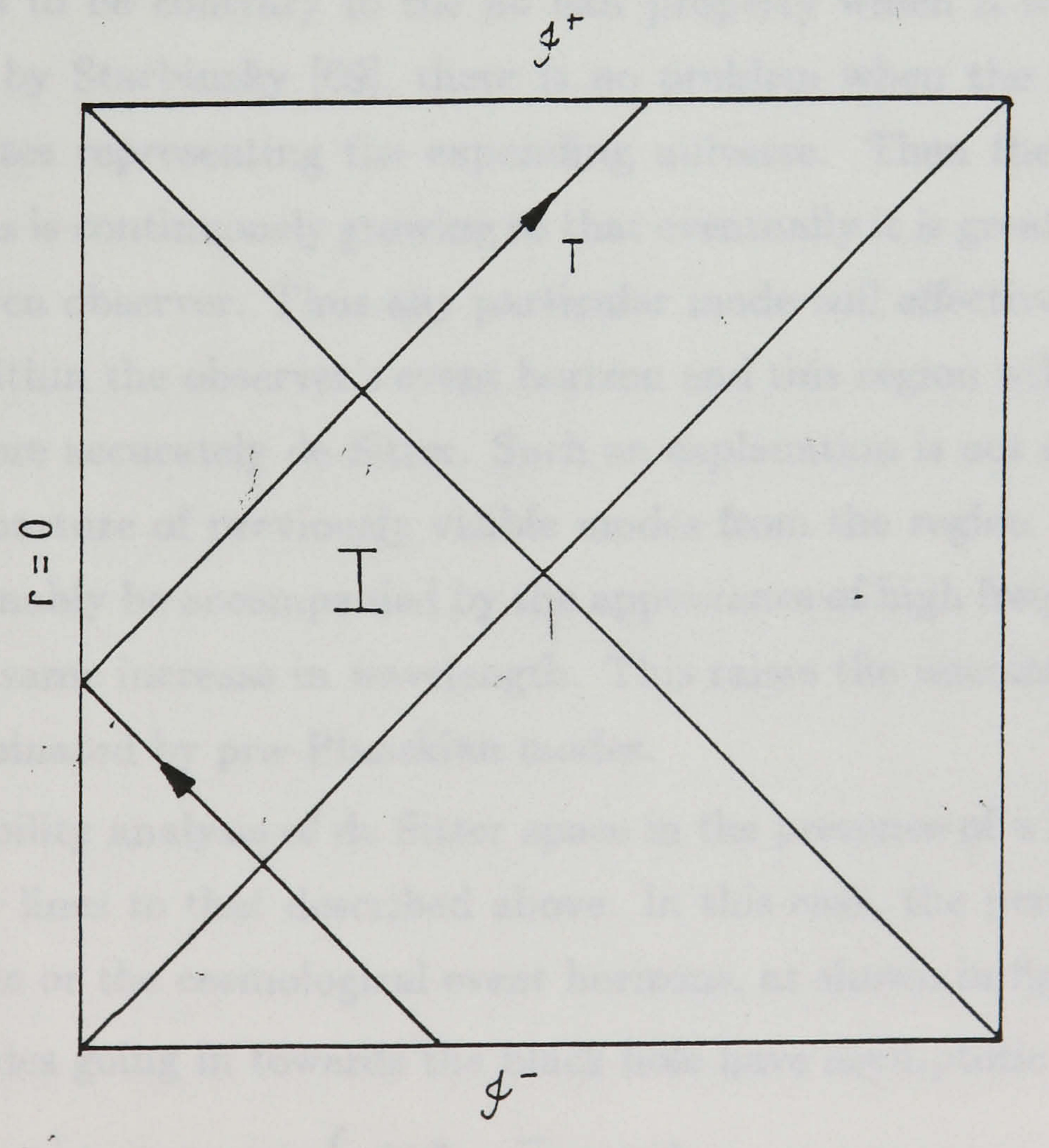

Fig.3.2 De Sitter space with the wave vector of a perturbation.

Asymptotically the wave is described by

$$Z_i \to e^{i\sigma r^*} + T_i e^{-i\sigma r^*} \ , \quad r^* \to \infty \ . \tag{3.4.8}$$

The solutions to (3.4.4) should satisfy this boundary condition. Likewise, the complex conjugates of these solutions, which are themselves solutions, should satisfy the complex conjugate boundary conditions

$$Z_i^* \to e^{-i\sigma r^*} + T_i^* e^{i\sigma r^*} \ , \quad r^* \to \infty \ . \tag{3.4.9}$$

The Wronskian of any two solutions, $f_1(x)$ and $f_2(x)$, and defined by

$$[f_1, f_2] = \frac{df_1}{dx} f_2 - f_1 \frac{df_2}{dx} \ , \tag{3.4.10}$$

is a constant since its derivative is zero. Thus the Wronskian at $r^* \to \infty$, where solutions take the form (3.4.8) should equal that at $r^* = 0$, where $Z = 0$. Thus $0 = TT^* - 1$, or $T = e^{i\delta_l}$, where δ_l is a phase shift. The perturbations therefore remain bounded – they do not blow up, but nor does their amplitude decrease. This appears to be contrary to the no hair property which is sought. However, as pointed out by Starbinsky [69], there is no problem when the metric is described in co–ordinates representing the expanding universe. Then the wavelength of the perturbations is continuously growing so that eventually it is greater than the horizon size for a given observer. Thus any particular mode will effectively cease to exist in the region within the observer's event horizon and this region will appear to become more and more accurately de Sitter. Such an explanation is not entirely satisfactory since the departure of previously visible modes from the region within the horizon, would presumably be accompanied by the appearance of high frequency modes which undergo the same increase in wavelength. This raises the uncomfortable notion of a universe dominated by pre–Planckian modes.

The stability analysis of de Sitter space in the presence of a black hole proceeds along similar lines to that described above. In this case, the perturbation can cross the black hole or the cosmological event horizons, as shown in figure 3.3.

The modes going in towards the black hole have asymptotic forms

$$\overleftarrow{Z}_i \to \begin{cases} e^{i\sigma r^*} + \overleftarrow{R}_i e^{-i\sigma r^*} \ , & r^* \to \infty \ , \\ \overleftarrow{T}_i e^{i\sigma r^*} \ , & r^* \to -\infty \ , \end{cases} \tag{3.4.11}$$

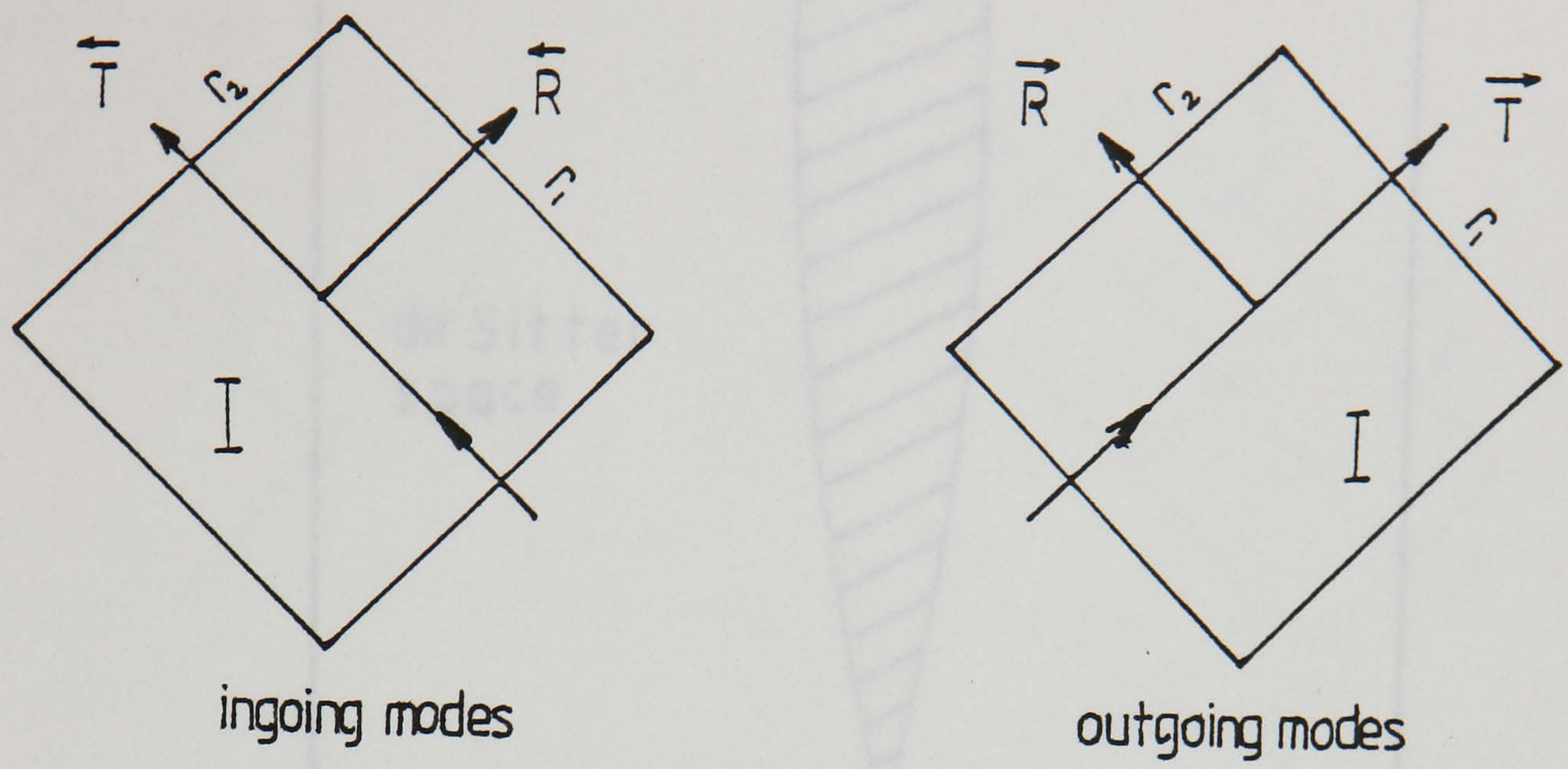

Fig.3.3 The modes $\overleftarrow{Z}$ and $\overrightarrow{Z}$ in region I.

with reflection and transmission coefficients $\overleftarrow{R}$ and $\overleftarrow{T}$. The outward going modes have the form

$$\overrightarrow{Z}_i \rightarrow \begin{cases} \overrightarrow{T}_i e^{i\sigma r^*} , & r^* \rightarrow \infty , \\ e^{-i\sigma r^*} + \overrightarrow{R}_i e^{i\sigma r^*} , & r^* \rightarrow -\infty . \end{cases} \tag{3.4.12}$$

This set of modes, with real σ, form a complete set. Furthermore, they remain bounded on the horizons and therefore a general perturbation, given its mode of expansion, also remains bounded.

This result on the stability of black holes in de Sitter spacetimes can be applied to the inflationary scenario. Consider a situation where there is a local enhancement of the radiation density sufficiently large to overcome the cosmological expansion. This density enhancement will collapse to form a black hole, as shown in figure 3.4. Outside, the metric will have the form of a perturbed black hole in de Sitter spacetime, and an observer will see the space approach de Sitter space. In the more general situation, there may be many such density enhancements on a variety of scales, leading to many black holes separated by regions which approach de Sitter space.

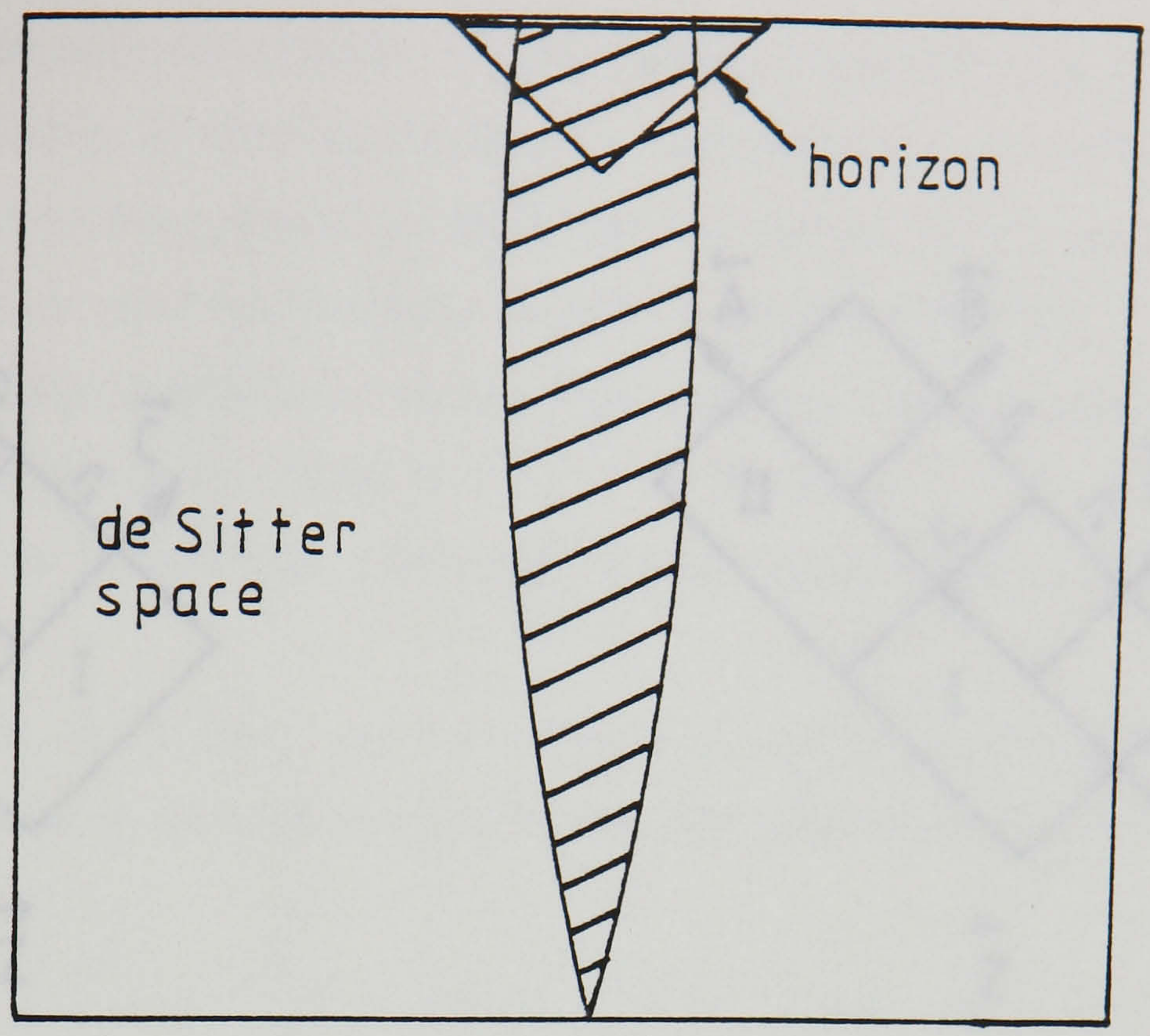

Fig.3.4 A large perturbation of de Sitter space which collapses to a singularity.

3.5 Stability of the Cauchy Horizon

The perturbation equations derived in §3.3 also enable one to examine the effect of the presence of a cosmological constant on the stability of the Cauchy horizon of a charged black hole. Initially, perturbations are set up outside the hole, in the region labelled I in figure 3.5. The relevant mode functions should therefore be set up either on the past event horizon, r_2, of the black hole or on the past cosmological event horizon, r_1. These modes represent ingoing modes $\overleftarrow{Z}$ and outgoing modes $\overrightarrow{Z}$ respectively. In both cases the modes have unit amplitude on the appropriate past horizon.

Near the Cauchy horizon, the modes have an asymptotic form

$$Z \to A(\sigma)e^{-i\sigma r^*} + B(\sigma)e^{+i\sigma r^*} \ , \quad r^* \to +\infty \ . \tag{3.5.1}$$

Specifically, I shall concentrate on the behaviour at the horizon AB. To distinguish this horizon from the other Cauchy horizon at AF it is necessary to restore the time

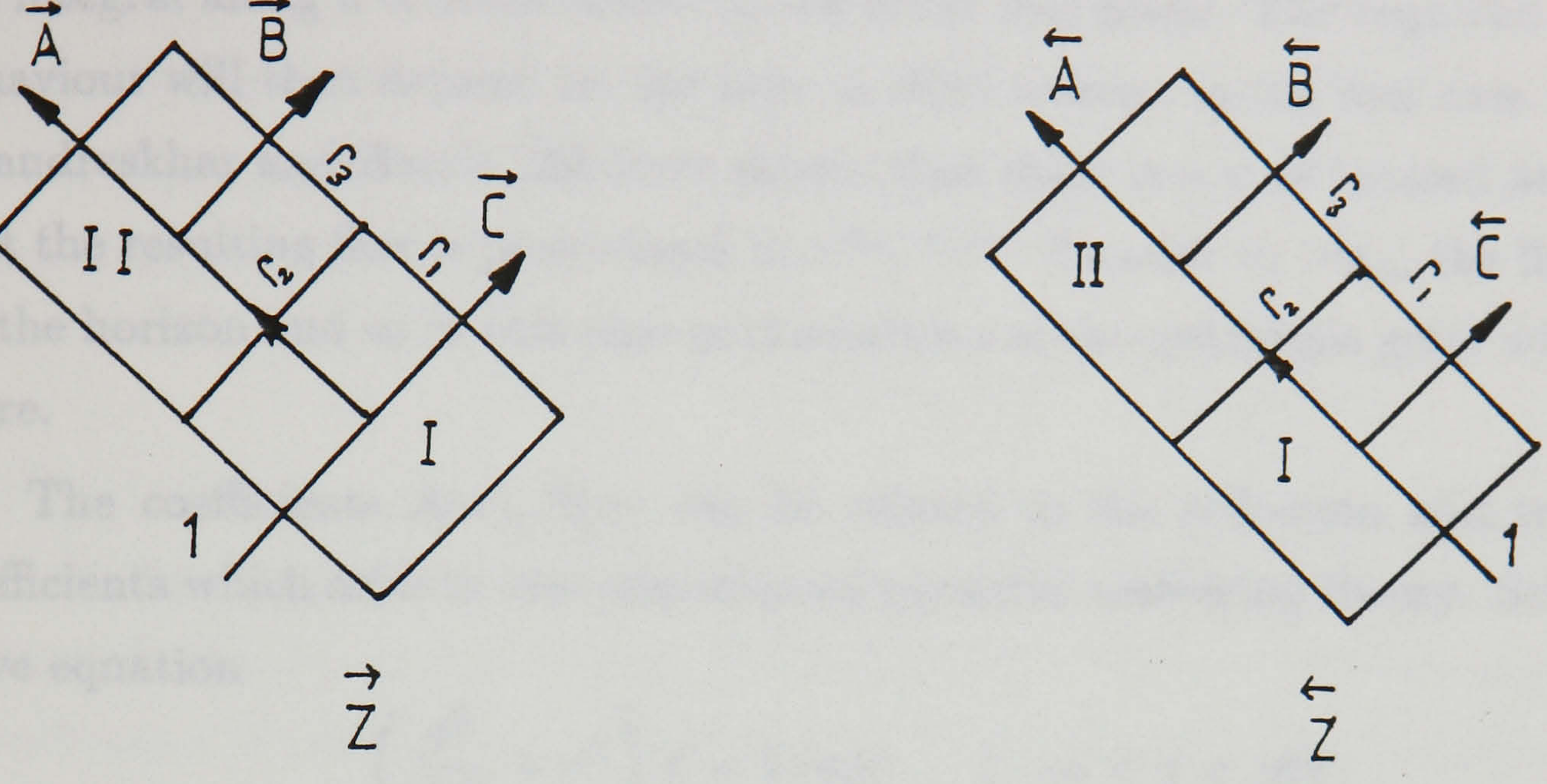

Fig.3.5 The modes $\overleftarrow{Z}$ and $\overrightarrow{Z}$ in regions I and II.

dependence, $e^{i\sigma t}$, of the solutions. The Fourier transform, $z(t)$, constructed from the mode functions Z and an initial amplitude W is

$$z(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W Z e^{i\sigma t} d\sigma\ . \tag{3.5.2}$$

Since in region II, $u = r^* + t$ and $v = r^* - t$, this can be written as

$$z(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W(\sigma)\left[A(\sigma)\, e^{-i\sigma v} + B(\sigma)\, e^{i\sigma u}\right] d\sigma\ . \tag{3.5.3}$$

At the Cauchy horizon AB, u remains finite whilst $v \to \infty$ (as distinguished from AF where v is finite whilst $u \to \infty$). The four velocity of any observer here approaches $\partial/\partial V$, where $V = -e^{-\kappa_3 v}$, and they see a flux which is basically given by $z^* \frac{\partial z}{\partial V} - z\frac{\partial z^*}{\partial V}$. Since z is bounded, it is necessary to concentrate on the behaviour of $\frac{\partial z}{\partial V}$. This is given by

$$\frac{\partial z}{\partial V} = \frac{1}{\kappa_3} e^{\kappa_3 v}\frac{\partial z}{\partial v} \tag{3.5.4}$$

$$= \frac{i e^{\kappa_3 v}}{2\pi\kappa_3}\int_{-\infty}^{\infty} W(\sigma)A(\sigma)\sigma e^{-i\sigma v} d\sigma\ . \tag{3.5.5}$$

Since the exponential term in the integral (3.5.5) decays for negative imaginary σ, the behaviour of the flux at the Cauchy horizon is therefore ascertained by evaluating the integral along a contour closed in the lower half plane. The required asymptotic behaviour will then depend on the pole in $A(\sigma)$ nearest to the real axis. For $\Lambda = 0$, Chandreskhar and Hartle [20] have shown that there is a pole located at $-i\kappa_2$, and that the resulting flux is proprtional to $e^{(\kappa_3-\kappa_2)v}$. Because $\kappa_3 > \kappa_2$, the flux diverges on the horizon and so in this case perturbations of the spacetime grow without limit there.

The coefficients $A(\sigma), B(\sigma)$ can be related to the reflection and transmission coefficients which arise in one–dimensional potential scattering theory. Schrödinger's wave equation

$$\left(\frac{d^2}{dx^2} + \sigma^2\right) f = V(x) f \;, \quad (-\infty < x < \infty) \;, \tag{3.5.6}$$

has solutions with asymptotic behaviours $e^{\pm i\sigma x}$ for $x \to \pm\infty$. In particular, there are two solutions with asymptotic behaviours:

$$\begin{aligned} f_1(x,\sigma) &\to e^{-i\sigma x} \quad && x \to \infty \\ f_2(x,\sigma) &\to e^{i\sigma x} \quad && x \to -\infty \;. \end{aligned} \tag{3.5.7}$$

The Wronskian $[f_1(x,\sigma), f_1(x,-\sigma)]$ shows that $f_1(x,\sigma)$ and $f_1(x,-\sigma)$ are independent solutions. Similarly, $f_2(x,\sigma)$ and $f_2(x,-\sigma)$ are also independent. It follows that there are expansions in which one solution can be constructed out of the others. i.e.

$$f_2(x,\sigma) = \frac{R_1(\sigma)}{T_1(\sigma)} f_1(x,\sigma) + \frac{1}{T_1(\sigma)} f_1(x,-\sigma) \tag{3.5.8}$$

and

$$f_1(x,\sigma) = \frac{R_2(\sigma)}{T_2(\sigma)} f_2(x,\sigma) + \frac{1}{T_2(\sigma)} f_2(x,-\sigma) \;, \tag{3.5.9}$$

or asymptotically

$$\lim_{x \to +\infty} f_2(x,\sigma) \to \frac{R_1(\sigma)}{T_1(\sigma)} e^{-i\sigma x} + \frac{1}{T_1(\sigma)} e^{i\sigma x} \tag{3.5.10}$$

and

$$\lim_{x \to -\infty} f_1(x,\sigma) \to \frac{R_2(\sigma)}{T_2(\sigma)} e^{i\sigma x} + \frac{1}{T_2(\sigma)} e^{-i\sigma x} \;. \tag{3.5.11}$$

Then $T_1(\sigma)f_2(x,\sigma)$ represents an incident wave of unit amplitude from $+\infty$, producing a reflected wave of amplitude $R_1(\sigma)$ and a transmitted wave of amplitude $T_1(\sigma)$. Likewise, $T_2(\sigma)f_1(x,\sigma)$ represents an incident wave of unit amplitude from $-\infty$, producing reflected and transmitted wave of amplitudes $R_2(\sigma)$ and $T_2(\sigma)$ respectively.

For the black hole in de Sitter space, the asymptotic behaviour in the interior region II is given by

$$Z \to \begin{cases} A(\sigma)e^{-i\sigma r^*} + B(\sigma)e^{+i\sigma r^*} & r^* \to +\infty \\ e^{-i\sigma r^*} & r^* \to -\infty\,. \end{cases} \tag{3.5.12}$$

This can be compared with the general expressions (3.5.7) and (3.5.10) for the behaviour of $f_2(x,\sigma)$ in these limits. The behaviour at $r^* \to -\infty$ takes the same form in both cases if one puts $\sigma \to -\sigma$ in (3.5.7). Then, making the same substitution in (3.5.10), one can deduce that

$$A_{II}(\sigma) = \frac{1}{T_{II}(-\sigma)} \tag{3.5.13}$$

and

$$B_{II}(\sigma) = \frac{R_{II}(-\sigma)}{T_{II}(-\sigma)}\,. \tag{3.5.14}$$

However, the behaviour of the modes at $r^* \to -\infty$ must be matched on either side of the future black hole event horizon BE. In order to do this one must change to Kruskal co–ordinates since v is well defined on this horizon whilst all the other co–ordinates tend to infinity†. Thus, restoring time dependence $e^{i\sigma t}$, the behaviour in region II at the outer horizon is

$$Z_{II} \to e^{i\sigma(t-r^*)} = e^{-i\sigma v} \quad r^* \to -\infty\,. \tag{3.5.15}$$

In region I, the ingoing and outgoing modes have different behaviours and must therefore be treated separately. Considering first the ingoing modes, $\overleftarrow{Z}$, (3.4.9) gives the asymptotic behaviour as

$$\overleftarrow{Z}_I \to \overleftarrow{T}_I(-\sigma)e^{i\sigma(t+r^*)} = \overleftarrow{T}_I(\sigma)e^{i\sigma v} \quad r^* \to -\infty\,, \tag{3.5.16}$$

† In region I v is given by $v = r^* + t$, whilst in region II it is given by $v = r^* - t$

or, putting $\sigma \to -\sigma$

$$\overleftarrow{Z}_I \to \overleftarrow{T}_I(-\sigma)e^{-i\sigma v} \quad r^* \to -\infty\,. \tag{3.5.17}$$

Then the modes $\overleftarrow{Z}_I$ and $\overleftarrow{Z}_{II}$ can be matched if the latter is rescaled by $\overleftarrow{T}_I(-\sigma)$. Thus the expression (3.5.13) for $A(\sigma)$ becomes

$$\overleftarrow{A}_{II}(\sigma) = \frac{\overleftarrow{T}_I(-\sigma)}{\overleftarrow{T}_{II}(-\sigma)}\,. \tag{3.5.18}$$

Similar reasoning applies in the case of the outgoing modes, $\vec{Z}$, where the behaviour in the exterior region is given by

$$\vec{Z}_{II} \to \vec{R}_I(\sigma)e^{i\sigma(t+r*)} = \vec{R}_I(\sigma)e^{i\sigma v} \quad r^* \to -\infty\,, \tag{3.5.19}$$

or

$$\vec{Z}_{II} \to \vec{R}_I(-\sigma)e^{-i\sigma v} \quad r^* \to -\infty\,. \tag{3.5.20}$$

In this case matching the modes gives

$$\vec{A}_{II}(\sigma) = \frac{\vec{R}_I(-\sigma)}{\overleftarrow{T}_{II}(-\sigma)}\,. \tag{3.5.21}$$

Equation (3.5.5), together with the relations (3.5.18) and (3.5.21), provide the behaviour of the flux of ingoing and outgoing modes as they reach the Cauchy horizon. If there are no poles in $A(\sigma)$ above $-i\kappa_3$ then the flux will not diverge – the horizon will appear stable to observers attemting to cross it. In the case of a Reissner–Nordström black hole in flat space, Chandrasekhar and Hartle [20] supposed that the coefficient $A(\sigma)$ is simply given by (3.5.13). $T^{-1}(-\sigma)$ was shown to have poles along the imaginary axis at multiples of $i\kappa_-$ and $i\kappa_+$, where κ_- and κ_+ are the surface gravities on the horizons on the incident and transmission sides of the potential barrier. In this case, it follows that the flux diverges; the horizon is unstable. However, the scattering of the perturbations off the spacetime external to the hole should be taken into account. For black holes in de Sitter space there is the possibility that the relevent pole in $T_{II}^{-1}(-\sigma)$ will cancel with a zero in $\overleftarrow{T}_I(-\sigma)$ and $\vec{R}_I(-\sigma)$. This is precisely what does happen and thus the Cauchy horizon for the de Sitter black hole is stable.

The analytic structure of $T^{-1}(-\sigma)$ can be obtained by the arguments of Chandreskhar and Hartle with virtually no modification. First note that equations (3.5.7), (3.5.10) and (3.5.11) imply that the transmission coefficients can be related to the Wronskian of f_1 and f_2.

$$\frac{1}{T_1(-\sigma)} = \frac{1}{T_2(-\sigma)} = \frac{1}{2i\sigma}\left[f_1(x,-\sigma), f_2(x,-\sigma)\right] . \tag{3.5.22}$$

Thus the construction of approximate solutions for f_1 and f_2 permits the derivation of an expression for T_{II}^{-1}. Let

$$f_2(x,-\sigma) = e^{-i\sigma x} + \Psi(x,\sigma) , \quad \text{with } \Psi \to 0 \text{ as } x \to -\infty . \tag{3.5.23}$$

From (3.5.6) Ψ satisfies the differential equation

$$\left(\frac{d^2}{dx^2} + \sigma^2\right)\Psi = \left(e^{-i\sigma x} + \Psi\right) V . \tag{3.5.24}$$

This can be solved by multiplying both sides of (3.5.24) by the Green function

$$G(x-x') = \frac{1}{\sigma}\mathrm{sin}\sigma(\mathrm{x}-\mathrm{x}')\,\theta(\mathrm{x}-\mathrm{x}') , \tag{3.5.25}$$

where $\theta(x-x')$ is the step function

$$\theta(x-x') = \begin{cases} 1 & \text{if } x > x' \\ 0 & \text{if } x < x'. \end{cases} \tag{3.5.26}$$

Integration then yields

$$\Psi(x,\sigma) = \frac{1}{\sigma}\int_{-\infty}^{x} \mathrm{sin}\sigma(x-x')V(x')\left(e^{-i\sigma x'} + \Psi\left(x',\sigma\right)\right) dx' . \tag{3.5.27}$$

Thus the solution $f_2(x,-\sigma)$ can be established through the iterative process defined in (3.5.27). An approximate solution can be obtained by noting that all the potentials of interest, i.e. (3.3.26) and (3.3.36), have the asymptotic form

$$V \to \begin{cases} c_- e^{-2\kappa_- x} & x \to \infty \\ c_+ e^{2\kappa_+ x} & x \to -\infty. \end{cases} \tag{3.5.28}$$

Using the form of $V(x)$ as $x \to -\infty$ and performing only the first iteration gives

$$f_2(x,-\sigma) \sim e^{-i\sigma x} + \frac{1}{4\kappa_+} \frac{c_+}{(\kappa_- - i\sigma)} e^{(2\kappa_+ - i\sigma)x} \ , \quad x \to -\infty \ . \tag{3.5.29}$$

A similar analysis as $x \to +\infty$ gives an approximate form of f_1:

$$f_1(x,-\sigma) \sim e^{i\sigma x} + \frac{1}{4\kappa_-} \frac{c_-}{(\kappa_- - i\sigma)} e^{-(2\kappa_- - i\sigma)x} \ , \quad x \to \infty \ . \tag{3.5.30}$$

A rough guide to the behaviour of the transmission coefficient is found by using the results (3.5.29) and (3.5.30) at $x = 0$. Then

$$\frac{1}{T(-\sigma)} \approx 1 - \frac{c_+}{4i\sigma\kappa_+} - \frac{c_-}{4i\sigma\kappa_-} + \frac{c_+ c_-}{16 i\sigma\kappa_+\kappa_-} \frac{(\kappa_+ + \kappa_- - i\sigma)}{(\kappa_+ - i\sigma)(\kappa_- - i\sigma)} \ . \tag{3.5.31}$$

Clearly, T^{-1} has poles at $\sigma = -i\kappa_\pm$. The more rigorous analysis of Chandrasekhar and Hartle shows that there is an infinite string of poles at $\sigma = -im\kappa_\pm$, $m =$ integer, along the imaginary σ axis. These poles are illustrated in figure 3.6.

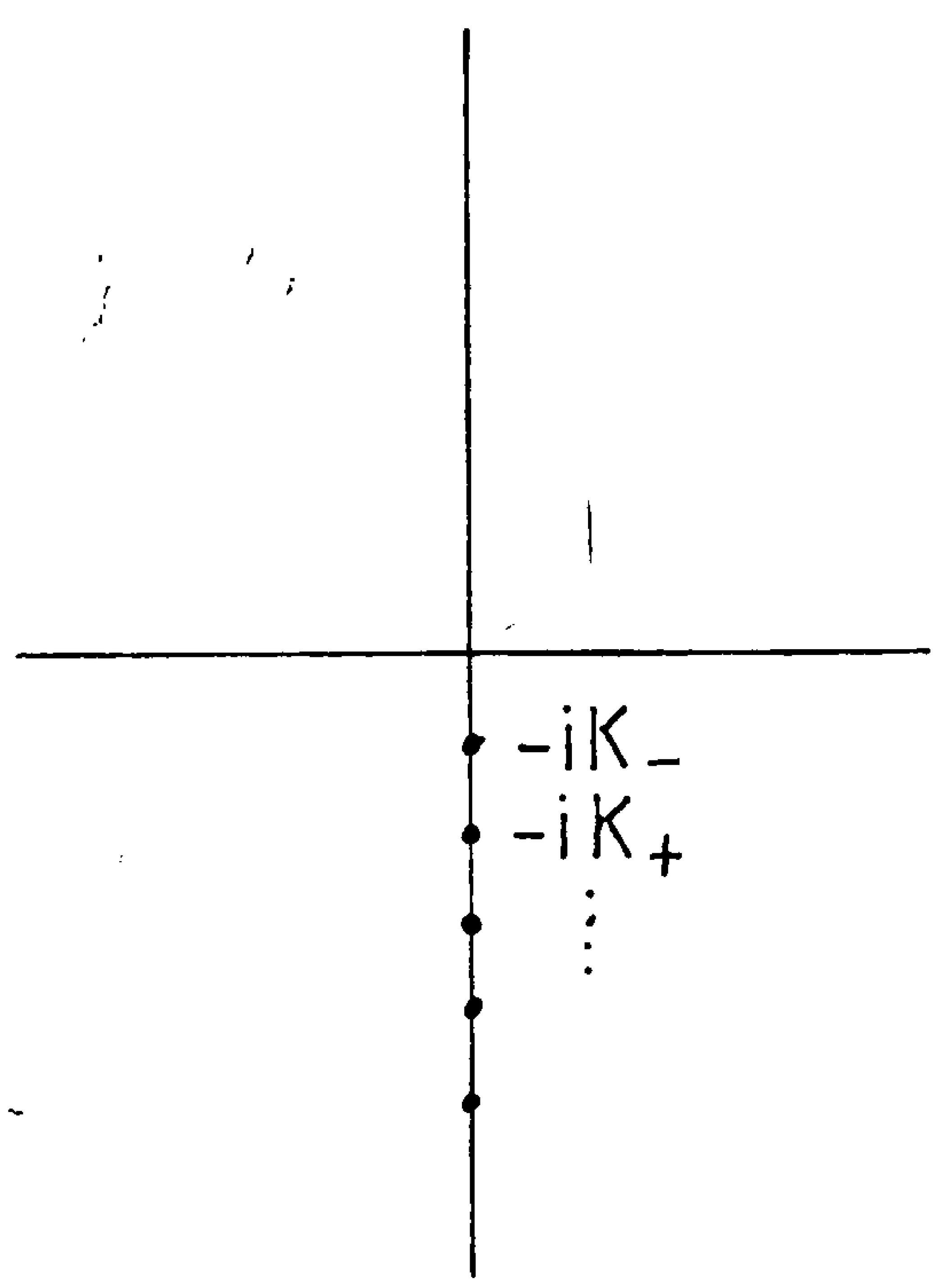

Fig.3.6 The poles in the transmission coefficient, $T_{II}^{-1}(-\sigma)$.

For the cases of interest, in region I $\kappa_\pm = (\kappa_1, \kappa_2)$ and in region II $\kappa_\pm = (\kappa_2, \kappa_3)$. Thus

$$\frac{T_I(i\kappa_2)}{T_{II}(i\kappa_2)} \approx \frac{c_3}{c_1}\frac{\kappa_1^2}{\kappa_3^2}\frac{(\kappa_2-\kappa_1)}{(\kappa_2-\kappa_3)}, \tag{3.5.32}$$

and the pole at $\sigma = -i\kappa_2$ in T_{II}^{-1} has cancelled with a zero in T_I. Similarly, there is a zero in R_I which also cancels the pole in T_{II}^{-1} at $-i\kappa_2$. Without this cancellation, the pole at $-i\kappa_2$ dominates the integral in (3.5.5) so that the flux goes as $e^{(\kappa_3-\kappa_2)v}$ and is therfore divergent. The cancellation of this pole means that the leading pole in $A(\sigma)$ from T_{II}^{-1} is now at $-i\kappa_3$. Thus the closure of the contour along the real axis in the lower half plane can approach infinitesimally close to $\sigma = -i\kappa_3$, and the term $e^{-i\sigma v}$ in (3.5.5) which formerly caused the divergence now cancels with the factor $e^{\kappa_3 \nu}$. Thus the pole results in a flux on the Cauchy horizon which is now finite.

The only other poles in $A(\sigma)$ come from poles in $\overleftarrow{T}_I(-\sigma)$ or $\overrightarrow{R}_I(-\sigma)$. Stability of the Cauchy horizon requires that neither of these amplitudes have poles in that region of the lower half plane in which $T_{II}^{-1}(-\sigma)$ is analytic. The existence of any such poles would mean that in region I the modes were dominated by those terms with coefficients $R_I(\sigma)$ and $T_I(\sigma)$. Thus poles at $0 < \mathrm{Im}\,\sigma < \kappa_3$ would lead to modes with growing exponential behaviour (c.f.(3.4.9)),

$$Z \to \begin{cases} e^{-i\sigma r^*} & r^* \to \infty \\ e^{i\sigma r^*} & r^* \to -\infty. \end{cases} \tag{3.5.33}$$

This kind of mode is an unbound state or a 'quasi-normal' mode [21]. Their complex frequencies dominate the late time behaviour of perturbations outside the hole in a similar way that pure tones dominate in the ringing of a bell. Although the perturbations of a black hole may be due to many causes – the incidence of gravitational waves, an object falling into the hole, aspherical collapse – the decay of the perturbation will eventually take on a form characteristic of the black hole and entirely independent of the initial conditions.

The frequencies of the first few quasi-normal modes have been tabulated in tables 1 and 2. In table 1, with $\alpha = 8M$, the values of σ all lie in the range $0 < \mathrm{Im}\,\sigma < \kappa_3$, and result in instabilities on the Cauchy horizon. With $\alpha = 20M$ on the other hand, the values of σ lie outside the range $0 < \mathrm{Im}\,\sigma < \kappa_3$ when $Q = M$ (more precisely, when $Q > .999M$). This numerical approach cannot show rigorously that all of the modes fall outside the range but it does show up a clear trend. There is strong evidence that the Cauchy horizon is stable in this case. Increasing α further has only a small effect upon the frequencies, suggesting stability for sufficiently large α.

In the limit that $\alpha \to \infty$ the spacetime outside the hole becomes flat. In this limit the cancellation between the zeros and poles in $A(\sigma)$ leave an increasingly large residue, and this results in a large flux at the Cauchy horizon. The amplification factor is given by (3.5.31) where, in the limit of large α, $c_1 \sim \alpha^{-2}$ and $c_3 \sim M^{-2}$. If α is the present cosmological scale, and the radiation entering the black hole is in the form of starlight, then the flash of radiation near the Cauchy horizon would be an obstacle to passing through the black hole. As $\alpha \to \infty$, the radiation flux diverges and the Cauchy horizon becomes unstable.

The possibility of there being a stable Cauchy horizon for some black holes raises a number of conceptual difficulties. An observer, who had crossed the Cauchy horizon, could follow a timelike path into another asymptotic region of the universe. Identification of these regions would allow for closed timelike loops and would produce problems with acausality. Additionally, this observer would be able to see the singularity of the spacetime without falling into it. This is contrary to the cosmic censorship conjecture upon which the physical relevance of black holes is founded.

α=8M

Z_1

Q/M	l=1	l=2	l=3	l=4	κ_3
0	0.191 0.068i	0.352 0.075i	0.502 0.072i	0.648 0.075i	∞
0.2	0.192 0.069i	0.357 0.072i	0.507 0.075i	0.658 0.073i	2400
0.4	0.203 0.071i	0.376 0.074i	0.534 0.073i	0.691 0.074i	131.5
0.6	0.222 0.075i	0.404 0.078i	0.581 0.076i	0.750 0.076i	20.00
0.8	0.266 0.088i	0.473 0.080i	0.663 0.080i	0.846 0.081i	3.771
1.0	0.373 0.073i	0.610 0.076i	0.835 0.076i	1.058 0.076i	0.1531

Z_2

Q/M	l=1	l=2	l=3	l=4
0	0.087 0.090i	0.285 0.069i	0.455 0.071i	0.616 0.071i
0.2	0.089 0.091i	0.286 0.069i	0.461 0.072i	0.622 0.073i
0.4	0.091 0.090i	0.294 0.070i	0.470 0.073i	0.635 0.074i
0.6	0.096 0.091i	0.305 0.073i	0.495 0.075i	0.669 0.075i
0.8	0.106 0.092i	0.327 0.076i	0.533 0.077i	0.725 0.080i
1.0	0.094 0.092i	0.371 0.098i	0.609 0.074i	0.834 0.076i

Table 1. The complex characteristic frequencies of the quasi-normal modes for Z_1 and Z_2 with $\alpha = 8M$. The values of the surface gravity, κ_3 on the Cauchy horizon are also indicated.

α=20M

Z_1

Q/M	l=1	l=2	l=3	l=4	κ_3
0	0.252	0.440	0.654	0.813	
	0.063i	0.064i	0.068i	0.068i	∞
0.2	0.255	0.445	0.662	0.822	
	0.064i	0.065i	0.069i	0.068i	2400
0.4	0.263	0.460	0.682	0.848	
	0.064i	0.066i	0.070i	0.070i	131.5
0.6	0.275	0.510	0.743	0.915	
	0.065i	0.069i	0.075i	0.072i	20.00
0.8	0.292	0.569	0.782	0.995	
	0.067i	0.072i	0.072i	0.076i	1.552
1.0	0.447	0.649	0.853	0.999	
	0.075i	0.079i	0.080i	0.079i	0.0548

Z_2

Q/M	l=1	l=2	l=3	l=4
0	0.099	0.355	0.605	0.803
	0.068i	0.068i	0.076i	0.070i
0.2	0.099	0.355	0.606	0.811
	0.068i	0.068i	0.076i	0.072i
0.4	0.100	0.357	0.610	0.834
	0.068i	0.069i	0.075i	0.075i
0.6	0.101	0.360	0.617	0.865
	0.069i	0.069i	0.074i	0.078i
0.8	0.102	0.403	0.630	0.901
	0.069i	0.072i	0.072i	0.080i
1.0	0.103	0.419	0.690	0.979
	0.069i	0.069i	0.064i	0.067i

Table 2. The complex characteristic frequencies of the quasi-normal modes for Z_1 and Z_2 with $\alpha = 20M$. The values of the surface gravity, κ_3 on the Cauchy horizon are also indicated.

APPENDIX
HARMONICS ON A TWO SPHERE

There are four tensor harmonics on a two–sphere Q_{ab} , P_{ab} , S^o_{ab} , S^e_{ab} which form a complete orthogonal set for the expansion of any symmetric second–rank tensor on S^2. These can be constructed from the scalar and vector harmonics as follows:

Scalar harmonics

$Q(\theta,\phi)$ are the scalar eigenfunctions of the Laplacian on S^2. i.e.

$$Q_{|a}{}^{|a} = -l(l+1)\,Q\ , \quad l = 0,1,2... \tag{A.1}$$

Vector harmonics

Transverse vector harmonics $S_a(\theta,\phi)$ are vector eigenfunctions of the Laplacian on S^2 which are transverse. i.e.

$$S_{a|c}{}^{|c} = -(l^2+l-1)\,S_a\ , \quad l = 1,2... \tag{A.2}$$

with

$$S_a{}^{|a} = 0\ . \tag{A.3}$$

These are classified as odd and even by their behaviour under parity transformations.

A third set of vector harmonics $P_a(\theta,\phi)$ can also be constructed from the scalar harmonics.

$$P_a = \frac{1}{l(l+1)} Q_{|a}\ , \quad l = 1,2.... \tag{A.4}$$

satisfying the eigenvalue equation

$$P_{a|c}{}^{|c} = -(l^2+l-1)\,P_a \tag{A.5}$$

and the condition

$$P_a{}^{|a} = -Q\ . \tag{A.6}$$

Tensor harmonics

Transverse traceless tensor harmonics $G_{ab}(\theta, \phi)$ are tensor eigenfunctions of the Laplacian operator on S^2 which are transverse and traceless. However, these can be shown to have zero degeneracy and therefore do not exist.

Traceless tensor harmonics S_{ab} can be constructed from both the odd and even transverse vector harmonics.

$$S_{ab} = S_{a|b} + S_{b|a} \tag{A.7}$$

satisfying the equations

$$S_{ab|c}{}^{|c} = -(l^2 + l - 4)\, S_{ab} \tag{A.8}$$

$$S_{ab}{}^{|b} = -(l^2 + l - 2)\, S_a \tag{A.9}$$

$$S_{ab}{}^{|ab} = 0 \tag{A.10}$$

$$S^a_a = 0 \tag{A.11}$$

Two more types of tensor harmonics can be constructed from the scalar harmonics.

$$Q_{ab} = \frac{1}{2}\bar{g}_{ab} Q \tag{A.12}$$

and

$$P_{ab} = \frac{1}{l(l+1)} Q_{|ab} + \frac{1}{2}\bar{g}_{ab} Q \tag{A.13}$$

satisfying the equations

$$P_{ab|c}{}^{|c} = -(l^2 + l - 4)\, P_{ab} \tag{A.14}$$

$$P_{ab}{}^{|b} = -\frac{1}{2}(l^2 + l - 2)\, P_a \tag{A.15}$$

$$P_{ab}{}^{|ab} = \frac{1}{2}(l^2 + l - 2)\, Q \tag{A.16}$$

$$P^a_a = 0 \tag{A.17}$$

Normalization

The scalar harmonics Q are normalized such that

$$\int d^2x\, \sqrt{{}^2\bar{g}}\, \frac{1}{2}\bar{g}_{ab} Q_n \frac{1}{2}\bar{g}^{ab} Q_{n'} = \delta_{nn'}\ . \tag{A.18}$$

Then

$$\int d^2x\,\sqrt{{}^2\bar{g}}\,P^a_{\ n}P_{a\,n'} = \frac{2}{l(l+1)}\delta_{nn'}\ . \tag{A.19}$$

The vector harmonics S_a are normalized such that

$$\int d^2x\,\sqrt{{}^2\bar{g}}\,S^a_{\ n}S_{a\,n'} = \delta_{nn'}\ . \tag{A.20}$$

The normalizations of the tensor harmonics are then:

$$\int d^2x\,\sqrt{{}^2\bar{g}}\,Q^{ab}{}_{n}Q_{ab\,n'} = \delta_{nn'}\ , \tag{A.21}$$

$$\int d^2x\,\sqrt{{}^2\bar{g}}\,P^{ab}{}_{n}P_{ab\,n'} = C^s\delta_{nn'} = \frac{(l^2+l-2)}{l(l+1)}\delta_{nn'}\ , \tag{A.22}$$

and

$$\int d^2x\,\sqrt{{}^2\bar{g}}\,S^{ab}{}_{n}S_{ab\,n'} = C^v\delta_{nn'} = 2(l^2+l-2)\delta_{nn'}\ . \tag{A.23}$$

REFERENCES

1. A.H.Guth, Phys. Rev. **D23** (1981) 347.
2. A.Vilenkin, Phys. Lett. **B117** (1982) 25.
3. A.Vilenkin, Phys. Rev. **D27** (1983) 2848.
4. S.W.Hawking, in 'Astrophysical Cosmology' ed. H.A.Brück, G.V.Coyne and M.S.Longair (Pontifica Academia Scientarium, Vatican City, 1982).
5. S.W.Hawking, Nucl. Phys. **B239** (1984) 257.
6. J.B.Hartle and S.W.Hawking, Phys. Rev. **D28** (1983) 2960.
7. W.H.Zurek, Phys. Rev. **D24** (1981) 1516.
8. W.H.Zurek, Phys. Rev. **D26** (1982) 1862.
9. Th.Kaluza, Sitzungsber. Preuss. Akad. Phys. Math. **K1** (1921) 996.
10. O.Klein, Z.Phys. **37** (1926) 895.
11. A.Salam and E.Sezgin, Phys. Lett. **147B** (1984) 47.
12. S.W.Hawking, Nature **248** (1974) 30.
13. S.Coleman, Phys. Rev. **D15** (1977) 2929.
14. L.Callan and S.Coleman, Phys. Rev. **D16** (1977) 1762.
15. S.W.Hawking and I.G.Moss, Phys. Lett. **110B** (1982) 35.
16. J.J.Halliwell and J.B.Hartle, Phys. Rev. **D41** (1990) 1815.
17. G.W.Gibbons and J.B.Hartle, 'Real Tunnelling Geometries and the Large–Scale Topology of the Universe' DAMPT Preprint (1990).
18. J.J.Halliwell and J.Louko, Phys. Rev. **D39** (1989) 2209.
19. S.Coleman, Nucl. Phys. **B310** (1988) 643.
20. S.Chandresekhar and J.B.Hartle, Proc. R. Soc. Lond. **A384** (1982) 301.
21. S.Chandresekhar, 'The Mathematical Theory of Black Holes' (Oxford University Press, Oxford, 1983).
22. R.Arnowitt, S.Deser and C.W.Misner in 'Gravitation: An Introduction to Current Research' ed. L.Witten (Wiley, New York, 1962).

23. A.Vilenkin, Phys. Rev. **D30** (1984) 509.

24. A.Vilenkin, Phys. Rev. **D37** (1988) 888.

25. J.Von Neumann, 'Mathematische Grundlagen der Quantenmechanik' (Springer, Berlin, 1932) [English translation by R.T.Beyer, 'Mathematical Foundations of Quantum Mechanics' (Princeton University Press, Princeton, 1955)].

26. H.Everett, Reviews of Modern Physics **29** (1957) 454.

27. F.London and E.Bauer, in 'Quantum Theory and Measurement' ed. J.A.Wheeler and W.H.Zurek (Princeton University Press, Princeton, 1983).

28. C.Kiefer, Class. Quantum Grav. **4** (1987) 1369.

29. J.J.Halliwell, Phys. Rev. **D39** (1989) 2912.

30. T.Fukuyama and M.Morikawa, Phys. Rev. **D39** (1989) 462.

31. C.Kiefer, Class. Quantum Grav. **6** (1989) 561, Phys. Lett. **139A** (1989) 201.

32. M.Morikawa, Phys. Rev. **D40** (1989) 4023.

33. H.Zeh, Phys. Lett. **116A** (1986) 9, Phys. Lett. **126A** (1988) 311.

34. B.De Witt, in 'Relativity, Groups and Topology at the 1963 Les Houches School' ed. B.De Witt and C.De Witt (gordon and Breach, New York, 1964).

35. E.Cremmer and J.Scherk, Nucl. Phys. **B108** (1976) 409, Nucl. Phys. **B118** (1977) 61.

36. E.Witten, Nucl. Phys. **B186** (1981) 412.

37. S.Randjbar–Daemi, A.Salam and J.Strathdee, Nucl. Phys. **B214** (1983) 491.

38. E.Kolb and R.Slansky, Phys. Lett. **135B** (1984) 378.

39. K.Maeda and H.Nishino, Phys. Lett. **154B** (1985) 358, Phys. Lett. **158B** (1985) 381.

40. J.J.Halliwell, Nucl. Phys. **B286** (1987) 729.

41. S.R.Lonsdale, Phys. Lett. **175B** (1986) 312.

42. J.J.Halliwell, Nucl. Phys. **B266** (1986) 228.

43. J.J.Halliwell and S.W.Hawking, Phys. Rev. **D31** (1985) 1777.

44. G. 't Hooft, Phys. Rev. Lett. **37** (1976) 8, Phys. Rev. **D14** (1976) 3432.

45. J.S.Langer, Ann. Phys. **41** (1967) 108.

46. D.Atkatz and H.Pagels, Phys. Rev. **D25** (1982) 2065.
47. G.W.Gibbons, S.W.Hawking and M.J.Perry, Nucl. Phys. **B138** (1978) 141.
48. S.W.Hawking, Nucl. Phys. **B144** (1978) 349.
49. J.A.Wheeler, in ref.34.
50. G.W.Gibbons and S.W.Hawking, Commun. Math. Phys. **66** (1979) 291.
51. T.Eguchi, P.B.Gilkey and A.J.Hanson, Physics Reports **66** (1980).
52. B.Carter, in 'Black Holes' ed. C.De witt and B.S.De Witt (Gordon and Breach, New York, 1973).
53. G.W.Gibbons and S.W.Hawking, Phys. Rev. **D15** (1977) 2738.
54. A.A.Starobinski, Sov. Phys. JETP **37** (1973) 28.
55. G.W.Gibbons, Commun. Math. Phys. **44** (1975) 245.
56. D.N.Page, Phys. Lett. **79B** (1978) 235.
57. S.W.Hawking, Phys. Lett. **150B** (1985) 339.
58. D.Klebanov, T.Banks and L.Susskind, Nucl. Phys. **B317** (1989) 665.
59. T.Regge and J.A.Wheeler, Phys. Rev. **108** (1957) 1063.
60. F.J.Zerilli, Phys. Rev. **D2** (1970) 2141.
61. S.W.Hawking, Commun. Math. Phys. **25** (1972) 152.
62. W.Israel, Phys. Rev. **164** (1967) 1176.
63. B.Carter, Phys. Rev. Lett. **26** (1971) 331.
64. R.M.Wald, Phys. Rev. **D28** (1983) 2118.
65. K.Maeda, in 'Proceedings of the Fifth Marcel Grossman Meeting' (World Scientific, Singapore, 1989).
66. B.C.Xanthopoulos, Proc. R. Soc. Lond. **A378** (1981) 61.
67. J.D.Barrow, in 'The Early Universe' ed. G.W.Gibbons, S.W.Hawking and S.T. Siklos (Cambridge University Press, Cambridge, 1983).
68. W.Boucher and G.W.Gibbons, ibid.
69. A.A.Starobinsky, JETP **64** (1973) 48.