Preface

The Second Samos Meeting on Cosmology, Geometry and Relativity organised by the Research Laboratory for Geometry, Dynamical Systems and Cosmology (GEO.DY.SY.C.) of the Department of Mathematics, the University of the Aegean, took place at the Doryssa Bay Hotel/Village at Pythagoreon, site of the ancient capital, on the island of Samos from August 31st to September 4th, 1998. The Meeting focused on mathematical and quantum aspects of relativity theory and cosmology. The Scientific Programme Committee consisted of Professors D. Christodoulou and G.W. Gibbons and Dr S. Cotsakis, and the Local Organizing Committee comprised Professors G. Flessas and N. Hadjisavvas and Dr S. Cotsakis. More than 70 participants from 18 countries attended.

The scientific programme included 9 plenary (one hour) talks, 3 'semi-plenary' (30 minute) talks and more than 30 contributed (20 minute) talks. There were no poster sessions. However, a feature of the meeting was an 'open-issues' session towards the end whereat participants were given the opportunity to announce and describe open problems in the field that they found interesting and important. The open-issues discussion was chaired by Professor Gibbons and we include a slightly edited version of it in this volume.

This volume contains the contributions of most of the invited talks as well as those of the semi-plenary talks. Unfortunately the manuscripts of the very interesting talks by John Barrow about 'Varying Constants', Ted Jacobson on 'Trans-Plankian Black Hole Models: Lattice and Superfluid' and Tom Ilmanen's lecture on 'The Inverse Mean Curvature Flow of the Einstein Evolution Equations Coupled to the Curvature' could not be included in this volume.

The meeting was sponsored by the following organizations: the University of the Aegean, the Ministry of Civilization, the Ministry of Education and Religion and the Ministry of the Aegean, the National Research and Technology Secretariat, EPEAEK (EU funded program), the Municipality of Pythagoreon, the Union of Municipalities of Samos, and the Prefecture of Samos. All this support is gratefully acknowledged.

We wish to thank all those individuals who helped to make this meeting possible. In particular we are deeply indebted to Professor P.G.L. Leach (Natal) who contributed a great deal in many aspects before, during and after the event. The heavy duty of being Secretary to the Meeting was carried out with great success by Ms Thea Vigli-Papadaki with help from Mrs Manto Katsiani.

We also wish to express our sincere thanks to the staff of Springer-Verlag for their enormous and expert help in shaping this volume and, more generally, for the true interest they show in the series of these Samos meetings.

Karlovassi, Greece *Spiros Cotsakis*
Cambridge, UK *Gary Gibbons*
October 1999

Contents

Global Wave Maps on Curved Space Times

Yvonne Choquet-Bruhat

Gravitation et Cosmologie Relativiste, t.22-12,
Université Paris VI, 75252 Paris, France

Introduction

Wave maps from a pseudoriemannian manifold of hyperbolic (lorentzian) signature (V, g) into a pseudoriemannian manifold are the generalisation of the usual wave equations for scalar functions on (V, g). They are the counterpart in hyperbolic signature of the harmonic mappings between properly riemannian manifolds.

The wave map equations are an interesting model of geometric origin for the mathematician, in local coordinates they look like a quasilinear quasidiagonal system of second order partial differential equations which satisfy the Christodoulou [17] and Klainerman [18] null condition. They also appear in various areas of physics (cf. Nutku 1974 [6], Misner 1978 [7]).

The first wave maps to be considered in physics were the σ-models, for instance the mapping from the Minkowski spacetime into the three sphere which models the classical dynamics of four meson fields linked by the relation:

$$\sum_{a=1}^{4} \mid f^a \mid^2 = 1\,.$$

Wave maps play an important role in general relativity, in general integration problem or in the construction of spacetimes with a spatial isometry group. Indeed:

1. The harmonic coordinates, used for a long time in various problems, express that the identity map from (U, g), U domain of a chart of the spacetime, into an open set of a pseudoeuclidean space is a wave map. Wave maps from a spacetime (V, g) into a pseudoriemannian manifold $(V, \hat{e})$, with $\hat{e}$ a given metric on V, gives a global harmonic gauge condition on (V, g).
2. The Einstein, or Einstein-Maxwell, equations for metrics possessing a one parameter spacelike isometry group can be written as a coupled system of a wave map equation from a manifold of dimension three and an elliptic, time dependent, system of partial differential equations on a two dimensional manifold, together with ordinary differential equations for the Teichmuller parameters (Moncrief 1986 [12], YCB and Moncrief 1995 [20]).

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 1–29, 2000.

The natural problem for wave maps is the Cauchy problem. It is a nonlinear problem, complicated by the fact that the unknown does not take their values in a vector space, but in a manifold. Gu Chaohao 1980 [9] has proven global existence of smooth wave maps form the 2-dimensional Minkowski spacetime into a complete riemannian manifold by using the Riemann method of characteristics. Ginibre and Velo 1982 [10] have proven a local in time existence theorem for wave maps from a Minkowski spacetime of arbitrary dimensions into the compact riemannian manifolds $O(N)$, $CP(N)$, or $GC(N,p)$ by semigroup methods. They prove global existence on 2-dimensional Minkowski spacetime. These local and global results have been extended to arbitrary regularly hyperbolic sources and complete riemannian targets in YCB 1987 [13], which proves also global existence for small data on $n+1$ dimensional Minkowski spacetime, $n \geq 3$ and odd, due to the null condition property. This last result has been proved to hold for $n = 2$ by YCB and Gu Chaohao 1989 [16], if the target is a symmetric space and for arbitrary n by YCB 1998c [24].

Global existence of weak solutions, without uniqueness, for large data in the case of $2 + 1$ dimensional Minkowski space has been proved by Muller and Struwe 1996 [22]. Counter examples to global existence on $3 + 1$ dimensional Minkowski space have been given by Shatah 1988 [14] and Shatah and Tahvildar-Zadeh 1995 [21].

This article is composed of two parts. In Part A we give a pedagogical introduction to wave maps together with a new proof of the local existence theorem. In Part B we prove a global existence theorem of wave maps in the expanding direction of an expanding universe.

A. General Properties

1 Definitions

Let u be a mapping between two smooth finite dimensional manifolds V and M:

$$u: \quad V \longrightarrow M.$$

Let (x^α), $\alpha = 0, 1, \ldots, n$, be local coordinates in an open set ω of the source manifold V supposed to be of dimension $n + 1$. Suppose ω sufficiently small for the mapping u to take its value in a coordinate chart (y^A), $A = 1, \ldots, d$ of the target manifold M supposed to be of dimension d. The mapping u is then represented in ω by d functions u^A of the $n + 1$ variables x^α

$$(x^\alpha) \quad \mapsto \quad y^A = u^A(x^\alpha).$$

The mapping u is said to be differentiable at $x \in \omega \subset V$ if the functions u^A are differentiable. The notion is coordinate independent if V and M are differentiable.

The gradient $\partial u(x)$ of the mapping u at x is an element of the tensor product of the cotangent space to V at x by the tangent space to M at $u(x)$:

$$\partial u(x) \in T^*_x V \otimes T_{u(x)} M.$$

The gradient itself, ∂u, is a section of the vector bundle E with base V and fiber $E_x \equiv T^*_x V \otimes T_{u(x)} M$ at x.

We now suppose that the manifolds V and M are endowed with pseudo-riemannian metrics denoted respectively by g and h. We endow the vector bundle E with a connexion whose coefficients acting in $T^*_x V$ are the coefficients of the riemannian connexion at x of the metric g while the coefficients acting in $T_{u(x)} M$ are the pull back by u of the connexion coefficients of the riemannian connexion at $u(x)$ of the metric h, we denote by ∇ the corresponding covariant differential. If f is an arbitrary section of E represented in a small enough open set ω of V by the $(n+1) \times d$ differentiable functions f^A_α of the $n+1$ coordinates x, then its covariant differential is represented in ω by the $(n+1)^2 \times d$ functions

$$\nabla_\alpha f^A_\beta(x) \equiv \partial_\alpha f^A_\beta(x) - \Gamma^\mu_{\alpha\beta}(x) f^A_\mu(x) + \partial_\alpha u^B(x) \Gamma^A_{BC}(u(x)) f^C_\beta(x),$$

where $\Gamma^\mu_{\alpha\beta}$ and Γ^A_{BC} denote respectively the components of the riemannian connections of g and h.

The covariant differential of a section f of E is a section of $T_* V \otimes$E, also a vector bundle over V.

Analogous formulas using the Leibniz rule for the derivation of tensor products give the covariant derivatives in local coordinates of sections of bundles over V with fiber $\otimes\ {}^p T^*_x$V$\otimes\ {}^q$T$_{u(x)} M$. In particular:

1. The covariant differential ∇g of the metric g, section of $\otimes^2 T^* V$, is zero by the definition of its riemannian connection. The field $h(u)$ defined by u and the metric h, section of the vector bundle over V with fiber $\otimes^2 T_{u(x)}$ at x, has also a zero covariant derivative ∇h, pull back by u of the riemannian covariant derivative of h.
2. Commutation of covariant derivatives gives the following useful generalisation of the Ricci identity

$$(\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha) f^A_\lambda = R_{\alpha\beta\lambda}{}^\mu(x) f^A_\mu(x) + \partial_\alpha u^C \partial_\beta u^B R_{CB}{}^A{}_D f^D_\mu.$$

2 Wave Maps. Cauchy Problem

From now on we will suppose that the source (V, g) is lorentzian, i.e. that the metric g is of hyperbolic signature, which we will take to be $(-, +, ..., +)$.

The following definition generalizes to mappings into a pseudoriemannian manifold the classical definition of a scalar valued wave equation on a lorentzian manifold.

Definition. A mapping u: $(V,g) \to (M,h)$ is called a *wave map* if the trace with respect to g of its second covariant derivative vanishes, i.e. if it satisfies the following second order partial differential equation, taking its values in TM:

$$g.\nabla^2 u = 0.$$

In local coordinates on V and M this equation is:

$$g^{\alpha\beta}\nabla_\alpha\partial_\beta u^A \equiv g^{\alpha\beta}(\partial^2_{\alpha\beta}u^A - \Gamma^\lambda_{\alpha\beta}\partial_\lambda u^A + \Gamma^A_{BC}(u)\partial_\alpha u^B\partial_\beta u^C = 0.$$

The wave map equation reads thus in local coordinates as a semilinear quasidiagonal system of second order partial differential equations for d scalar functions u^A. The diagonal principal term is just the usual wave operator of the metric g; the nonlinear terms are a quadratic form in ∂u, with coefficients functions of u.

The wave map equation is invariant under isometries of (V,g) and (M,h): let u be a wave map from (V,g) into (M,h), let f and F be diffeomorphisms of V and M respectively, then $F\circ u\circ f$ is a wave map from $(f^{-1}(V), f_*g)$ into $(F(M), dFh)$.

Throughout this paper we stipulate that the manifold V is then of the type $S\times R$, with each submanifold $S_t \equiv S\times\{t\}$ space like. We denote by (x,t) a point of V.

Remark. If the source (V,g) is globally hyperbolic, i.e. the set of timelike paths joining two points is compact in the set of paths (Leray 1953 [1]), then it is isometric to a product $S\times R$ with each submanifold $S_t\equiv S\times\{t\}$ spacelike and a Cauchy surface, i.e. such that each timelike or null path without end point cuts S_t once (Geroch 1970 [4]).

The first natural problem to solve for a wave map is the *Cauchy problem*, i.e. the construction of a wave map taking together with its first derivative given values on a spacelike submanifold of V for instance S_0. The Cauchy data are a mapping φ from S into M and a section ψ of the vector bundle with base S and fiber $T_{\varphi(x)}$ over x, namely:

$$u(0,x) = \varphi(x)\in M, \quad \partial_t u(0,x) = \psi(x)\in T_{\varphi(x)}M.$$

The results known for Leray hyperbolic systems cannot be used trivially when the target M is not a vector space. However, the standard local in time existence and uniqueness results known for scalar-valued systems can be used to solve the local in time problem for wave maps by glueing local in space results (cf CB 1998a [23]). This local in time existence can also be deduced from those known from scalar valued systems by first embedding the target (M,h) into a pseudoriemannian manifold (Q,q) with Q diffeomorphic to R^n. We give here a variant of the obtention of a system of R^N valued partial differential equations equivalent, modulo hypothesis on the Cauchy data, to the wave map equation.

Lemma 1. Let u: $V\to M$ and i: $M\to Q$ be arbitrary smooth maps between pseudoriemannian manifolds (V,g), (M,h), (Q,q). Set $U\equiv i\circ u$, map from

(V, g) into (Q, q). Denote by ∇ the covariant derivative corresponding to the map on which it acts,then the following identity holds:

$$\nabla \partial U \equiv \partial i . \nabla \partial u + \nabla \partial i . (\partial u \otimes \partial u),$$

that is, if (x^α), (x^A) and (x^a) are respectively local coordinates on V, M and Q while ∇ is the covariant derivative either for the maps u: $(V, g) \to (M, h)$, or i: $(M, h) \to (Q, q)$ or U: $(V, g) \to (Q, q)$,

$$\nabla_\alpha \partial_\beta U^a \equiv \partial_A i^a \nabla_\alpha \partial_\beta u^A + \partial_\alpha u^A \partial_\beta u^B \nabla_A \partial_B i^a.$$

Proof. By the definition of the covariant derivative we have

$$\nabla_\alpha \partial_\beta U^a \equiv \partial^2_{\alpha\beta} U^a - \Gamma^\lambda_{\alpha\beta} \partial_\lambda U^a + \Gamma^a_{bc} \partial_\alpha U^b \partial_\beta U^c,$$

where Γ^a_{bc} are the coefficients of the riemannian connexion of (Q, q),

By the law of the derivation of a composition map we find

$$\partial_\alpha U^a \equiv \partial_\alpha (i \circ u)^a \equiv \partial_A i^a \partial_\alpha u^A,$$

$$\partial^2_{\alpha\beta} U^a \equiv \partial_A i^a \partial^2_{\alpha\beta} u^A + \partial^2_{AB} i^a \partial_\alpha u^A \partial_\beta u^B.$$

The given formula results from these expressions after adding and substracting the term $\partial_A i^a \Gamma^A_{BC} \partial_\alpha u^B \partial_\alpha u^C$ (up to names of summation indices). We obtain as announced:

$$\nabla_\alpha \partial_\beta U^a \equiv \partial_A i^a (\partial^2_{\alpha\beta} u^A - \Gamma^\lambda_{\alpha\beta} \partial_\lambda u^A + \Gamma^A_{BC} \partial_\alpha u^B \partial_\beta u^C)$$
$$+ (\partial^2_{AB} i^a - \Gamma^C_{AB} \partial_C i^A + \Gamma^a_{bc} \partial_A i^b \partial_B i^c) \partial_\alpha u^A \partial_\beta u^B. \quad (1)$$

Lemma 2. Suppose (M, h) is isometrically embedded in (Q, q), i.e. $h \equiv i_* q$, then $\nabla \partial i \in \otimes^2 T_* M \otimes TQ$ is the pull back on M of the second fundamental form K of $i(M)$ as submanifold of Q, it takes its values at a point $y \in i(M)$ in the subspace of $T_y Q$ orthogonal to $T_y i(M)$. We have in arbitrary coordinates on M and Q:

$$\nabla_A \partial_B i^a \equiv \partial_A i^b \partial_B i^c K^a_{cb}.$$

Proof. It is a classical result (cf. for instance [15, V 2, p 280]); it can be proved and explained as follows in adapted local coordinates of M and Q. Let (y^A), $A = 1, ..., d$ be local coordinates in the neighbourhood of a point $y_0 \in M$. We choose in a neighbourhood in Q of the point $i(y_0)$ local coordinates (z^a), $a = 1, ..., D$, such that the embedding i is represented in this neighbourhood by:

$$i^a(y) = y^a \quad \text{if} \quad a = 1, ..., d \quad \text{and} \quad i^a(y) = 0 \quad \text{if} \quad a = d + 1, ..., D.$$

We choose a moving frame with d axes such that $\theta^a = dy^a$, $a = 1, ..., d$, while the other $D - d$ axes are orthogonal to these ones and between themselves. In the neighbourhood considered the metric q of Q is then i:

$$q = \sum_{a,b=1}^{d} q_{ab} dy^a dy^b + \sum_{a=d+1}^{D} (\theta^a)^2,$$

The gradient of the mapping i: $(M,h) \to (Q,q)$ in the chosen coordinates and frame is:

$$\partial_A i^a = \delta^a_A, \quad A = 1,...,d \quad ; \quad a = 1,...,D.$$

Denote by Q^a_{bc} the connection coefficients of the metric q in the considered coframe, the covariant derivative of the gradient of a mapping i: $(M,h) \to (Q,q)$ is:

$$\nabla_B \partial_A i^a \equiv \partial^2_{BA} i^a - \Gamma^C_{BA} \partial_C i^a + Q^a_{bc} \partial_B i^b \partial_A i^c$$

which gives here:

$$\nabla_B \partial_A i^a = -\Gamma^a_{BA} + Q^a_{BA}, \quad \text{if} \quad a = 1,...,d;$$
$$\nabla_B \partial_A i^a = Q^a_{BA} \quad \text{if} \quad a = d+1,...,D.$$

If i is an isometric embedding we have on M that $q_{ab} = h_{ab}$, $a,b = 1,...,d$. We have then on $i(M)$ identified with M:

$$\Gamma^a_{bc} = Q^a_{bc}, \quad a,b,c = 1,...,d,$$

while Q^a_{BA}, $B,A = 1,...,d$; $a = d+1,...,D$ are the components of the pull back by i of the second fundamental form of $i(M)$ as submanifold on M, equal in the chosen coordinates' frame to the components K^a_{bc} of that form in the chosen frame orthogonal to the tangent space to $i(M)$.
Remark. Denote by $\nu^{(a)}$, $a = d+1,...,D$, the unit mutually orthogonal vectors orthogonal to i(M). In the chosen coordinates and frame the components of $\nu^{(a)}$ are

$$\nu^{(a)}_c = \delta^a_c \text{if} \quad a,c = d+1,\ldots,D, \quad \nu^{(a)}_c = 0 \text{if} c = 1,...,d.$$

We find therefore in this frame

$$\nabla_b \nu^{(a)}_c = -Q^a_{bc}, \quad b,c = 1,\ldots,d; \quad a = d+1,\ldots,D.$$

which gives the usual tensorial form for the components of the second fundamental form of $i(M)$ as an element of $\otimes^2 T_* i(M) \otimes TQ$.
Lemma 3. If the mapping u: $(V,g) \to (M,h)$ is a wave map and if the mapping i: $(M,h) \to (Q,q)$ is an isometric embedding then the mapping $U \equiv i \circ u$: $(V,g) \to (Q,q)$ satisfies in the considered local coordinates the following semilinear second order equation:

$$g^{\alpha\beta}\{\nabla_\alpha \partial_\beta U^a - \partial_\alpha U^c \partial_\beta U^b K^a_{bc}(U)\} = 0.$$

Proof. The proof results from lemmas 1 and 2 together with the fact that $\partial_\alpha U^a \equiv \partial_A i^a \partial_\alpha u^A$.

Suppose that the manifold (M,h) is properly riemannian and has a non-zero injectivity radius. Embed it isometrically in a riemannian space (Q,q)

such that $i(M)$ admits a tubular neighbourhood Ω in Q (geodesics orthogonal to $i(M)$ have a length bounded away from zero in this neighbourhood). The subset $\Omega \subset Q$ can be covered by domains of local coordinates of the previously considered type with $K(U)$ depending smoothly on U in Ω.

The system satisfied by U: $(V, g) \to (\Omega, q)$ is invariant by change of coordinates on M and $\Omega \subset Q$. We can write it intrinsically under the form, with $K(U)$ defined when $U \in \Omega$:

$$\{\nabla^2 U - K(U).(\partial U \otimes \partial U)\} = 0,$$

where the first dot is a contraction in g and the second dot a contraction in q.

Choose Q diffeomorphic to R^N, as it is always possible (Whitney theorem), then there exists global coordinates z^I on Q. In these coordinates the equation satisfied by the mapping U: $(V, g) \to (Q, q)$ reads as a system of second order semilinear system of partial differential equations for a set of scalar functions U^I, defined if $U \equiv \{U^I\} \in \Omega$.

If (M, h) is properly riemannian it is always possible (Nash theorem) to embed it isometrically in a euclidean space (R^N, e). If M is compact then $i(M)$ always admit a tubular neighbourhood Ω.
Remark. If q is a flat metric, the operator $g.\nabla^2 U$ reads as a linear operator, the usual wave operator on (V, g) for a set of scalar functions, when the coordinates z^I are the cartesian ones, the nonlinearities are concentrated in the term with coefficient K.

3 Local Existence Theorem. Global Problem

We will use the classical local existence theorem for Leray hyperbolic system applied to the system we have obtained for U by embedding (M, h) for instance in a euclidean space.

We first recall some definitions. We denote by greek letters spacetime indices while tensors on S are indexed with latin letters. A metric $\mathbf{g}$ on $V \equiv S \times R$ is written in boldface, a t dependent metric on S is denoted g_t or (g_{ij}). We write as usual the spacetime metric $\mathbf{g}$ in a moving frame with time axis at the point (x, t) orthogonal to S_t under the form:

$$\mathbf{g} = -N^2 dt^2 + g_{ij}\theta^i\theta^j, \quad \text{with} \quad \theta^i \equiv dx^i + \beta^i dt.$$

The function N, called lapse, is strictly positive; the vector β is called the shift; the induced metric on each S_t, $g_t \equiv g_{ij}dx^i dx^j$, is properly riemannian.
Definition 1. Let $I \equiv [t_0, L)$ be an interval of R. The hyperbolic metric $\mathbf{g}$ on $V \equiv S \times I$ is said to be regularly hyperbolic if:

(i) There exist positive and continuous functions of t, B_1 and B_2, such that for each $t \in I$ it holds on S that

$$0 < B_1 \leq N \leq B_2.$$

(ii) The metrics g_t, $t \in I$, induced by $\mathbf{g}$ on S_t, are equivalent to a given riemannian metric s on S,, that is, there exist positive and continuous functions of $t \in I$, A_1 and A_2, such that for any vector field ξ on S and $t \in I$ it holds on S that

$$A_1 s(\xi, \xi) \leq g_t(\xi, \xi) \leq A_2 s(\xi, \xi).$$

We suppose that the metric s has a non zero injectivity radius, hence is complete. The same property is then enjoyed by each g_t and the manifold $(V, \mathbf{g})$ is globally hyperbolic (CB 1967 [3]).

We now define functional spaces of tensor fields on S.

Definition 2. The Sobolev space W_s^p of tensors of some given type on S is the completion of the space of such tensors in C_0^∞ (i.e. infinitely differentiable with compact support on S) in the following norm:

$$\| f \|_{W_s^p} \equiv \{\sum_{k=0}^{p} \int_S | D^k f |^p \mu_s\}^{1/p},$$

where D denotes the covariant derivative, $|\ |$ the pointwise norm and μ_e the volume element in the metric e. We set $W_s^2 = H_s$.

With the given definition of the spaces W_s^p and the hypothesis that e has a nonzero injectivity radius the usual imbedding and multiplication properties of Sobolev spaces on R^n hold.

Our spaces W_s^p coincide with spaces of tensor fields whose generalised covariant derivatives in the metric e of order less or equal to p are in $L^p(\mu_s)$ if in addition to previous hypothesis we suppose that the curvature of the metric s is uniformly bounded as well as its derivatives of relevant order (cf. Aubin 1982 [11]).

Remark. For the local existence theorem the hypothesis that s has a nonzero injectivity radius can be replaced by its Sobolev regularity, i.e. by the hypothesis that the Sobolev embedding and multiplication properties hold: it is the case when (S, s) is a bounded open set of R^n enjoying the cone property (cf. for instance C.B-D.M [15, V 2, p 379]).

We now define functional spaces for tensor fields on V, noting first that a tensor of order P on V can be decomposed into a finite number of t-dependent tensors of order $\leq P$ on S. We say that the restriction to some given t of a tensor f on V belongs to a given functional space on S if it is so for each tensor of the above decomposition.

For simplicity of writing we take in this section the initial submanifold to be $t_0 = 0$.

Definition 3. We denote by $E_s^p(T)$ the Banach space of tensor fields on $V_T \equiv S \times [0, T]$ defined by

$$E_s^p(T) \equiv \mathrm{C}^k([0, T], W_{s-k}^p), \quad 0 \leq \mathrm{k} \leq \mathrm{s}.$$

We denote by E^p_s a space of tensors on V which are in $E^p_s(T)$ for any finite T. We set $E_s \equiv E^2_s$

Embedding and multiplication properties of the spaces $E^p_s(T)$ are an immediate consequence of these properties for the spaces W^p_s.

Theorem. Let $(V \equiv S \times I, \mathbf{g})$ be a regularly hyperbolic manifold with $[0,T] \subset I$. Let (M,h) be a smooth complete riemannian manifold embedded by i in an euclidean space R^N with cartesian coordinates z^I. Suppose that $D\mathbf{g}, \partial_t \mathbf{g} \in E_{s-1}(T)$.

Let φ, ψ be Cauchy data on S for a wave map $(V, \mathbf{g}) \to (M,h)$. Suppose that the corresponding set of functions $\Phi^I = (i \circ \varphi)^I$ and $\Psi^I = \partial i^I \psi$ on S are such that

$$\Phi^I \in H_s \text{and} \Psi^I \in H_{s-1}$$

Then if $s \geq \frac{n}{2} + 1$ there exists $\ell \ > 0$ and a wave map u taking the given data, and such that $U^I \equiv (i \circ u)^I \in E_s(\ell) \cap \Omega$.

The interval ℓ of existence for any s is equal to the interval corresponding to $s = s_0$, smallest integer greater than $\frac{n}{2} + 1$.

The solution is unique and depends continuously on the data. A solution with $U \in E_{s_0}(\ell)$, can be approximated by solutions with U in $E_s(\ell)$.

In the case $n = 2$ or 3 the result holds for $s_0 = 2$.

Proof. The existence and properties of $U \equiv (U^I)$ is classical (Leray theory, as completed by Dionne 1962 [2], YCB 1971 [5], YCB-Christodoulou-Francaviglia 1979 [8], one uses the fact that E_{s-1} is an algebra when $s - 1 > \frac{n}{2}$. The extension to $s = 2$ in the case $n = 2$ or 3 has been proved on Minkowski spacetime by Klainerman and Machedon [18], on curved spacetimes by Sogge 1993 [19](Fourier method) and C-B 1998a [23](energy estimates).

To show that U defines a wave map u taking the given Cauchy data we return to our adapted coordinates y^a in $\Omega \subset Q \equiv R^N$. If there exists a mapping u: $V \to M$ such that $U \ = i \circ u$, i.e. if U takes its values in $i(M)$, then we have the identity

$$g^{\alpha\beta}\{\nabla_\alpha \partial_\beta U^a - \partial_\alpha u^A \partial_\beta u^B \nabla_A \partial_B i^a\} \equiv \partial_A i^a g^{\alpha\beta} \nabla_\alpha \partial_\beta u^A.$$

The mapping U: $V \to i(M) \subset \Omega \subset Q$ annuls the left hand side, the right hand side is then also zero and u is a wave map taking the given Cauchy data. We thus have only to prove that U takes its values in $i(M)$, i.e. that $U^a = 0$ for $a = d+1, \ldots, D$. The equation satisfied by U reads:

$$g^{\alpha\beta}\{\partial^2_{\alpha\beta} U^a - \Gamma^\lambda_{\alpha\beta} \partial_\lambda U^a + \partial_\alpha U^b \partial_\beta U^c (Q^a_{bc} \text{ - } K^a_{bc})$$

with $K^a_{bc} = Q^a_{bc}$ if $a = d+1, \ldots, D$ (note also that $K^a_{bc} = 0$ for $a = 1, \ldots, d$), hence the $D - d$ functions U^a, $a = d+1, \ldots, D$, satisfy a linear homogeneous system, with zero Cauchy data by hypothesis. This system is only local, as are the coordinates y^a, however it is not difficult to deduce from it that U takes its values in $i(M)$ by using a partition of unity and the finite propagation speed of solutions of the wave equation.

Corollary. The theorem can be extended to local spaces, i.e. by replacing the spaces W^p_s on S by spaces of functions which are in W^p_s in each open relatively compact set $\omega_{(i)}$ of some locally finite covering of S, with uniformly bounded $W^p_s(\omega_{(i)})$ norms (cf. C-B 1998a).

Remark. It is possible to prove an analogous theorem with variants on the hypothesis on the metric $\mathbf{g}$. For instance less time regularity or (and) replacement of the spaces H_s on S by spaces W^p_s. One obtains eventually less time regularity of the solution.

The hypothesis made on the metric imply in all cases that $D\mathbf{g}$ is uniformly bounded on V_T. They do not necessarily imply that it is lipshitzian: the geodesics between two nearby points may not be unique.

Global existence lemma. Let $(V \equiv S \times [0,\infty),\mathbf{g})$ be a regularly hyperbolic manifold with $D\mathbf{g}$, $\partial_{\ t}\mathbf{g} \in E_s$, $s \geq s_0$. The wave map u with Cauchy data φ, ψ such that $(\Phi,\Psi) \in H_s \times H_{s-1}$. Then u exists globally on V if the norms $\| U(t,.) \|_{H_s}$ and $\| U(t,.) \|_{H_{s-1}}$ do not blow up in a finite time, i.e. are bounded by functions of t continuous on the interval $I \equiv [0,\infty)$.

Proof. It is a standard consequence of the local existence theorem, with the continuous dependence of the interval of existence on the $H_{s_0} \times H_{s_0-1}$ norm of the data.

In the next sections we will endeavour to estimate the involved $H_s \times H_{s-1}$ norms

4 First Energy Estimate

To study global problems for wave maps one must use their special geometric properties, as for other fundamental equations of physics.

The first quantity of physical significance is the energy of the map. In contradistinction with the case, where the source is riemannian, the energy of the map is not the spacetime Dirichlet integral (which is not a positive quantity in the lorentzian case) but a space integral analogous to the energy associated with a solution of the wave equation. We introduce it now.

The *stress energy tensor* of a mapping u: $(V,g) \to (M,h)$ is the covariant 2-tensor on V given by:

$$\mathrm{T(u)} \equiv (\mathrm{h}\circ\mathrm{u})(\partial\mathrm{u},\partial\mathrm{u}) - \tfrac{1}{2}\ \mathrm{g}\{\mathrm{g}\otimes(\mathrm{h}\circ\mathrm{u})\}.\{\partial\mathrm{u}\otimes\partial\mathrm{u}\}$$

that is

$$T_{\alpha\beta} = h_{AB}(u)\partial_\alpha u^A \partial_\beta u^B - \tfrac{1}{2} g_{\alpha\beta} g^{\lambda\mu} h_{AB}(u)\partial_\lambda u^A \partial_\mu u^B$$

which we will usually write:

$$T_{\alpha\beta} \equiv \partial_\alpha u.\partial_\beta \mathrm{u} - \tfrac{1}{2} g_{\alpha\beta}\partial_\lambda u.\partial^\lambda \mathrm{u}.$$

Indices are raised with $\mathbf{g}$, a dot denotes the scalar product in the metric h of the target space.

Lemma 1. The stress energy tensor $T(u)$ of a wave map u has zero divergence.

Proof. The metrics $\mathbf{g}$ and h have zero covariant derivative, therefore

$$\nabla_\alpha T^\alpha_\lambda \equiv \partial_\lambda u . g^{\alpha\beta} \nabla_\alpha \partial_\beta u \equiv h_{AB}(u) \partial_\lambda u^A g^{\alpha\beta} \nabla_\alpha \partial_\beta u^B = 0$$

if u is a wave map.

Corollary. The stress energy tensor of the mapping $U \equiv i \circ u$: $(V, \mathbf{g}) \to (R^N, q)$, i an isometric embedding of (M, h) into (R^N, q) has zero divergence if u is a wave map.

Proof. If (M, h) is isometrically embedded by i in (R^N, q) then the stress energy tensors of u and $U \equiv i \circ u$ are the same tensors on V, as can be seen by elementary calculus.

The *energy momentum vector* of the mapping u, equivalently of $U = i \circ u$, with respect to a vector X on V is the vector $\mathcal{P}(X, u)$ on V given in local coordinates by

$$\mathcal{P}^\alpha \equiv T^\alpha_\beta X^\beta$$

Lemma 2. If X is time like or null, then $\mathcal{P}$(X,u) is time like or null, X and $\mathcal{P}(X, u)$ have opposite time orientation.

Proof. Straightforward, cf. CB 1998a [23].

Lemma 3. The divergence of the energy momentum vector $\mathcal{P}(X, u)$ is given by

$$\nabla_\alpha \mathcal{P}^\alpha = \tfrac{1}{2} T^{\alpha\beta} (L_X \mathbf{g})_{\alpha\beta}, \qquad (L_X \mathbf{g})_{\alpha\beta} \equiv \nabla_\alpha X_\beta + \nabla_\beta X_\alpha.$$

The energy momentum vector $\mathcal{P}$ has zero divergence if X is a Killing vector of $\mathbf{g}$.

Proof. Straightforward, using the fact that the stress energy tensor has zero divergence. The symmetric 2-tensor $\pi \equiv L_X \mathbf{g}$ is the Lie derivative of the spacetime metric with respect to X.

The energy density of a mapping u at time t with respect to the past oriented timelike or null vector X is the non negative number

$$\epsilon(X, \nu) \equiv \mathcal{P}^{\ \alpha} \nu_\alpha$$

with $\mathcal{P}^{\ \alpha}$ the components of the energy momentum vector $\mathcal{P}(X, u)$ of u with respect to X and ν_α the components of the past oriented unit normal ν to S_t.

The mappings u and $U = i \circ u$ have the same energy density if i is an isometric embedding.

In the coframe $\theta^{\ \alpha}$ we have

$$\nu_i = 0, \quad \nu_0 = N$$

hence

$$\mathcal{P}^\alpha \nu_\alpha = \mathcal{P}^0 N.$$

If the space time metric $\mathbf{g}$ is stationary, i.e. admits a time like Killing vector, it is appropiate to define the energy density with respect to this vector. Otherwise the natural geometric choice is to take for X the past oriented unit normal ν to S_t. The energy momentum vector is then $\mathcal{P}$(ν,ν) and one obtains the usual energy density of u (equivalently of $U \equiv i \circ$ u), denoted $\epsilon(u)$, namely:

$$\mathcal{P}\ {}^0 N \equiv \epsilon(u) \equiv T^{00} N^2 \equiv \tfrac{1}{2}(\mid N^{-1}\partial_0 u \mid_h^2 + \mid Du \mid_{g,h}^2)$$

Also, if i is an isometric embedding in a euclidean space (R^N, δ) and $U \equiv i\circ$u,

$$\epsilon(u) \equiv \tfrac{1}{2}(\mid N^{-1}\partial_0 U \mid^2 + \mid DU \mid_g^2) \equiv \tfrac{1}{2}\delta_{IJ}\{g^{ij}\partial_i U^I \partial_j U^J + N^{-2}\partial_0 U^I \partial_0 U^J\},$$

We have denoted by $\mid\ \mid_{g,h}$ (respectively $\mid\ \mid_g$) the norm both in g and h (respectively in g and δ).

The integral of the energy density of u on S_t is, by definition, the *energy* $e(t,u)$ of u *at time* t. We denote by μ_t the volume element of g_t, we have:

$$e(t,u) \equiv \int_{S_t} \mathcal{P}^0 N \mu_t.$$

We deduce from the hypothesis that $\mathbf{g}$ is uniformly equivalent to the given metric e on S that $|\mathrm{Du}|_{g,h}^2 \equiv \mid DU \mid_g^2$ is uniformly equivalent to $\mid DU \mid_e^2 \equiv \mid DU \mid^2$. We see that $e(t,u)$ is uniformly equivalent to a sum of norms defined in Sect. 3: there exist positive numbers $C_{\mathbf{g}}$ and $C'_{\mathbf{g}}$ depending only on the bounds on g and N such that:

$$C_{\mathbf{g}} e(t,u) \leq \parallel \partial_0 u(.,t) \parallel_{L^2} + \parallel Du(.,t) \parallel_{L^2} \leq C'_{\mathbf{g}} e(t,u)$$

We denote by K the extrinsic curvature of S imbedded in $(V, \mathbf{g})$. In local coordinates (t, x^i) we have

$$K_{ij} = -\frac{1}{2N}(\partial_t g_{ij} + \nabla_i \beta_j + \nabla_j \beta_i)$$

We will prove the following theorem.

Theorem 1. (energy equality). Let u be a solution of the wave map equation on a manifold $V = \mathrm{S}\times I$ with a C^1 regularly hyperbolic metric $\mathbf{g}$ such that DN and NK are uniformly bounded in g norm on each S_t. Suppose that $u \in C^2(T) \cap E_1(T)$. Then u satisfies for $t \in I \equiv [0,T]$ the fundamental energy inequality:

$$e(t,u) = e(0,u) + \int_0^t \int_{S_\tau} N^{-1}\partial_i N \partial^i u.\partial_0 u + NK^{ij}T_{ij}\}\mu_\tau d\tau$$

Proof. A straightforward computation shows that for $X = \nu$ we have:

$$(L_X \mathbf{g})_{0i} = -\partial_i N, \qquad (L_X \mathbf{g})_{00} = 0, \qquad (L_X \mathbf{g})_{ij} = -2\ \omega\ {}_{ij}^{0} N = K_{ij}$$

The integration of the divergence equation satisfied by $\mathcal{P}$, the value of $L_\nu \mathbf{g}$ and the density of $C_0^\infty(S)$ in H_1 give the theorem.

In cosmological problems it is often convenient to take as time parameter the mean extrinsic curvature of the submanifolds S_t, which characterises the expansion (or contraction) of the universe. We set:

$$\tau \equiv Tr_g K \equiv g^{ij} K_{ij}$$

We will deduce from the energy equality the following corollary.

Corollary. We set

$$P_{ij} \equiv K_{ij} - \tfrac{1}{n} g_{ij}\tau, \quad \text{with} \quad Tr_g P \equiv g^{ij} P_{ij} = 0.$$

Then

$$e(t,u) = e(0,u) + \int_0^t \int_{S_s} N^{-1}\partial_i N \partial^i u.\partial_0 u + NP^{ij}\partial_i u.\partial_j u\}$$
$$+ N\tau\{(\tfrac{1}{n} - \tfrac{1}{2})\{\mid Du \mid_{g,h}^2 + \tfrac{1}{2} \mid N^{-1}\partial_0 u \mid_h^2\}\mu_s ds.$$

Proof. We have:

$$K^{ij}T_{ij} \equiv \{P^{ij} + \frac{1}{n}g^{ij}\tau\}\{\partial_i u.\partial_j u - \frac{g_{ij}}{2}(-N^{-2}\partial_0 u.\partial_0 u + \mid Du \mid^2_{g,h}\}$$

that is

$$K^{ij}T_{ij} \equiv P^{ij}\partial_i u.\partial_j u + \tau\{(\frac{1}{n} - \frac{1}{2})\{\mid Du \mid^2_{g,h} + \frac{1}{2} \mid N^{-1}\partial_0 u \mid^2_h\}$$

Theorem 2. (General energy inequality). Under the hypothesis of Theorem 1 the energy of a wave map satisfies the following inequality:

$$e(t,u) \leq e(0,u)exp\{\int_0^t Sup_{S_\tau}(\mid DN \mid_g + C \mid NK \mid_g)d\tau.$$

with C a positive number depending only on n.

Proof. The integral equality of Theorem 1 together with the inequality satisfied by scalar products imply the following inequality, with C a positive number depending only on n:

$$e(t,u) \leq e(0,u) + \int_0^t \int_{S_\tau}\{\mid DN \mid_g \mid Du \mid_{g,h} \mid N^{-1}{}_{\partial_0} u \mid_h + C \mid NK \mid_g (\mid Du \mid^2_{g,h} + \mid N^{-1}\partial_0 u \mid^2_h\}\mu_\tau d\tau$$

hence

$$e(t,u) \leq e(0,u) + \int_0^t Sup_{S_\tau}(\mid DN \mid_g + C \mid NK \mid_g)e_\tau(u)d\tau.$$

This inequality implies the theorem by the Gromwall lemma.

Remark 1. In the case where X is a Killing vector field of **g** and we use it to define the energy density the energy inequality becomes an equality, expressing the conservation of energy of the mapping u. We have chosen here for X the unit normal to S. It is a Killing field if $DN = 0$ and $K = 0$ the corresponding energy $e(t,u)$ is then conserved .

Remark 2. The energy inequality gives only an estimate of ∂U. An estimate of U, as a mapping in R^N, can be obtained from its initial data by the formula

$$U^I(.,t) = U^I(.,0) + \int_0^t \partial_t U^I(.,\tau)d\tau$$

which implies

$$\parallel U(.,t) \parallel_{L^2} \leq \parallel U(.,0) \parallel_{L^2} + t^{1/2} \parallel \partial_t U \parallel_{L^2}.$$

We will return later to the exploitation of the corollary of Theorem 1.

5 Second Energy Estimate

The estimate of the L^2 norm of Du and $\partial_0 u$ on S_t is not sufficient to prove the existence of strong solutions of the wave map eqation even for $n = 1$.

We will now obtain a local in time estimate of the H_1 norms of these quantities by a new method which will be better suited for the cosmological problems. We suppose the shift to be zero, then $\partial_0 \equiv \partial/\partial t$. We denote by $\bar{\nabla}$ the covariant derivative for mappings between the riemannian manifolds $(S,g) \to (M,h)$, acting on sections of vector bundles $\bar{E}^{(p,q)}$ over S with fiber $\otimes T^*_x \otimes^q T_{u(x)}M$, for example:

$$\bar{\nabla}_i \partial_j u^A \equiv \partial^2_{ij}u^A - \Gamma^h_{ij}\partial_h u^A + \Gamma^A_{BC}(u)\partial_i u^B \partial_j u^C.$$

We set (suggestion due to V. Moncrief):

$$e^{(1)}(t,u) \equiv \frac{1}{2}\int_{S_t}\{\bar{\Delta}u.\bar{\Delta}u + \mid \bar{\nabla}u' \mid^2_{g,h}\}\mu_t, \quad \text{with} \quad \text{u'} \equiv N^{-1}\partial_0 u.$$

where $\bar{\Delta}$ is the laplace operator for the metric g and the derivative $\bar{\nabla}$, i.e.:

$$\bar{\Delta}u \equiv g^{ij}\bar{\nabla}_i\partial_j u$$

We denote by D_t the covariant derivative of a mapping from R into time-dependent sections of a vector bundle $\bar{E}$, defined by:

$$D_t\partial_i u^A \equiv \partial_0\partial_i u^A + \Gamma^A_{BC}(u)\partial_0 u^B\partial_i u^C.$$

D_t is a linear operator mapping the space of time-dependent sections of $\bar{E}^{(p)}$ into itself given by the law

$$D_t\bar{\nabla}^p u^A \equiv \partial_0\bar{\nabla}^p u^A + \Gamma^A_{BC}(u)\partial_0 u^B\bar{\nabla}^p u^C.$$

The $\bar{\nabla}$ or D_t derivatives of the mappings from S or R into $\otimes\ {}^2TM$ by $x \mapsto h(u(x,t))$ or $t \mapsto h(u(.,t)$ are both zero. The D_t derivative of the metric g_{ij} is equal to $\partial_0 g_{ij} = -2NK_{ij}$. The following commutation relations can be foreseen and checked by straightforward computation:

$$D_t\partial_i u = \bar{\nabla}_i\partial_0 u,$$

$$D_t\bar{\nabla}_i\partial_0 u^A\text{-}\ \bar{\nabla}_i D_t\partial_0 u^A = R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E,$$

$$D_t\bar{\nabla}_i\partial_j u^A\text{-}\ \bar{\nabla}_i D_t\partial_j u^A = R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_j u^E\ \text{-}\ \partial_h u^A\partial_0\Gamma^h_{ij}.$$

We recall the identities (zero shift)

$$\partial_0 g^{ij} = 2NK^{ij}.$$

$$\partial_0\Gamma^h_{ij} \equiv \bar{\nabla}^h(NK_{ij}) - \bar{\nabla}_i(NK^h_j) - \bar{\nabla}_j(NK^h_i).$$

from which we deduce

$$D_t\bar{\nabla}^i\partial_i u = \bar{\nabla}^i D_t\partial_i u + R_{CD\ E}\partial_0 u^C\partial^i u^D\partial_i u^E + 2NK^{ij}\bar{\nabla}_j\partial_i u$$
$$+\{-\bar{\nabla}^h(N\tau) + 2\bar{\nabla}_i(NK^{ih})\}\partial_h u.$$

Before computing the time derivative of $e^{(1)}$ we set:

$$I_0 = \frac{1}{2}\mid \bar{\nabla}u' \mid^2_{g,h}, \quad I_1 \equiv \frac{1}{2}\bar{\Delta}u.\bar{\Delta}u,$$

hence

$$e^{(1)}(t,u) \equiv \int_{S_t}\{I_0 + I_1\}\mu_t.$$

We have

$$\frac{\partial\mu_t}{\partial t} = -N\tau$$

hence

$$\frac{de^{(1)}}{dt} = \int_{S_t}\{\frac{\partial}{\partial t}(I_0 + I_1) - N\tau(I_0 + I_1)\}\mu_t$$

We find, using the definition of D_t, the Leibnitz rule and the property $D_t h = 0$

$$\partial_0 I_1 \equiv D_t I_1 = D_t\bar{\nabla}^i\partial_i u.\bar{\nabla}^j\partial_j u,$$

We have

$$D_t\bar{\nabla}^i\partial_i u = g^{ih}D_t\bar{\nabla}_h\partial_i u + \partial_0 g^{ih}\bar{\nabla}_h\partial_i u$$

with

$$\partial_0 g^{ih} = 2NK^{ih}.$$

Using the commutation formulas and Stokes' formula we obtain:

$$\int_{S_t} D_t I_1 \mu_t = \int_{S_t} \{-D_t\partial_i u.\bar{\nabla}^i\bar{\nabla}^j\partial_j u + \mathrm{F}.\bar{\nabla}^j\partial_j u\}\mu_t$$

with

$$F \equiv R_{CD\ E}\partial_0 u^C\partial^i u^D\partial_i u^E + 2NK^{ij}\bar{\nabla}_j\partial_i u + \{-\bar{\nabla}^h(N\tau) + \mathbf{2}\bar{\nabla}_i(NK^{ih})\}\partial_h u.$$

On the other hand

$$\partial_0 I_0 \equiv D_t I_0 = g^{ij} D_t\bar{\nabla}_i u'.\bar{\nabla}_j u' + \tfrac{1}{2}\partial_0 g^{ij}\bar{\nabla}_i u'.\bar{\nabla}_j u'.$$

therefore:

$$\int_{S_t} D_t I_0 \mu_t = \int_{S_t} \{g^{ij} D_t\bar{\nabla}_i u'.\bar{\nabla}_j u' + NK^{ij}\bar{\nabla}_i u'.\bar{\nabla}_j u'\}\mu_t.$$

We compute the wave map equation $N^{-2}\nabla_0\partial_0 u^A - g^{ij}\nabla_i\partial_j u^A = 0$ with our definitions. We have, with $\omega^{\alpha}_{\beta\gamma}$ the connection coefficients of $\mathbf{g}$

$$\nabla_0\partial_0 u^A \equiv \partial_0\partial_0 u^A - \omega^{\alpha}_{00}\partial_\alpha u^A + \Gamma^A_{BC}\partial_0 u^C\partial_0 u^D$$

which gives:

$$\begin{aligned}\nabla_0\partial_0 u^A &\equiv N\partial_0(N^{-1}\partial_0 u^A) - N\partial^i N\partial_i u^A + \Gamma^A_{BC}\partial_0 u^C\partial_0 u^D\\ &\equiv N\{D_t(N^{-1}\partial_0 u^A - \partial^i N\partial_i u^A\}.\end{aligned}$$

On the other hand

$$\nabla_i\partial_j u = \bar{\nabla}_i\partial_j u - \omega^0_{ij}\partial_0 u = \bar{\nabla}_i\partial_j u + N^{-1}K_{ij}$$

The wave map equation reads therefore

$$D_t(N^{-1}\partial_0 u^A) = \bar{\nabla}^i(N\partial_i u^A) + \tau\partial_0 u^A.$$

The commutation relation written for $\partial_0 u$ applies to $u' \equiv N^{-1}\partial_0 u$, we have

$$(D_t\bar{\nabla}_i - \bar{\nabla}_i D_t)(N^{-1}\partial_0 u^A) = N^{-1}R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E,$$

We have therefore if u is a wave map

$$D_t\bar{\nabla}_i(N^{-1}\partial_0 u^A) = \bar{\nabla}_i\{\bar{\nabla}^j(N\partial_j u^A) + \tau\partial_0 u^A\} + N^{-1}R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E.$$

Inserting this expression in $D_t I_0$, adding $D_t I_1$ and integrating we find:

$$\begin{aligned}&\int_{S_t} D_t(I_0 + I_1)\mu_t =\\ &\int_{S_t}\{\bar{\nabla}_i(\bar{\nabla}^j(N\partial_j u) + \tau\partial_0 u) + N^{-1}R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E\\ &+ NK_{ij}\bar{\nabla}^j u'\}.\bar{\nabla}^i u'\mu_t + \int_{S_t}\{-D_t\partial_i u.\bar{\nabla}^i\bar{\nabla}^j\partial_j u + F.\bar{\nabla}^j\partial_j u\}\mu_t.\end{aligned}$$

The derivatives of third order in u cancel if u is a wave map. Indeed a straightforward computation (recall that $u' \equiv N^{-1}\partial_0 u$ and $D_t\partial_i u = \bar{\nabla}_i\partial_0 u$) gives

$$\bar{\nabla}^i(\bar{\nabla}^j(N\partial_j u).\bar{\nabla}_i u' - \bar{\nabla}^i\bar{\nabla}^j\partial_j u.\bar{\nabla}_i\partial_0 u \equiv C$$

with, by elementary computation,

$$C \equiv - \bar{\nabla}^i\bar{\Delta} u\partial_i N.u' + (\partial_i N\bar{\Delta} u + \partial^j N\bar{\nabla}_i\partial_j u).\bar{\nabla}^i u'$$

Under integration on S_t this term is equivalent to the following one, denoted B:

$$B \equiv \bar{\Delta} N\bar{\Delta} u.u' + (2\partial_i N\bar{\Delta} u + \partial^j N\bar{\nabla}_i\partial_j u).\bar{\nabla}^i u'$$

We have found:

$$\frac{de^{(1)}}{dt} = \int_{S_t} \{\partial_0(I_0+I_1) - N\tau(I_0+I_1)\}\mu_t$$
$$= \int_{S_t}\{\ A_1 +\ A_0 + B - N\tau\ (I_0+I_1)\mu_t$$

with

$$A_0 \equiv \{\bar\nabla_i(N\tau u') + N^{-1}R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E\ +\ NK_{ij}\bar\nabla^j u'\}.\bar\nabla^i u'$$

Setting as in the previous section

$$K_{ij} \equiv P_{ij} + \tfrac{1}{n}g_{ij}\tau, \quad \text{with} \quad g^{ij}P_{ij} = 0$$

gives

$$A_0 \equiv 2N(1+\tfrac{1}{n})\tau\ I_0 + NP_{ij}\bar\nabla^i u'.\bar\nabla^j u'\ +$$
$$\{\partial_i(N\tau)u' + N^{-1}R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E\ \}.\bar\nabla^i u'$$

while $A_1 \equiv F.\bar\nabla^j\partial_j u$, given by

$$A_1 \equiv \{R_{CD\ E}\partial_0 u^C\partial^i u^D\partial_i u^E + 2NK^{ij}\bar\nabla_j\partial_i u$$
$$+(-\bar\nabla^h(N\tau) + \mathbf{2}\bar\nabla_i(NK^{ih}))\partial_h u\}.\bar\Delta u$$

can be written:

$$A_1 \equiv N\tau\tfrac{4}{n}I_1 + 2NP^{ij}\bar\nabla_i\partial_j u.\bar\Delta u + \{R_{CD\ E}\partial_0 u^C\partial^i u^D\partial_i u^E$$
$$+[-\bar\nabla^h(N\tau) + \mathbf{2}\bar\nabla_i(NK^{ih})]\partial_h u\}.\bar\Delta u$$

We obtain the following theorem by summing and rearranging the various terms that we have found.

Theorem. (second energy equality). If u is a wave map, its second energy satisfies the following equality.

$$\frac{de^{(1)}(t,u)}{dt} = \int_{S_t}\{\ \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}\}\ \mu_t$$

with

$$\mathrm{I} \equiv N\tau\ [(1+\tfrac{\mathbf{2}}{n})\ I_0\ +\ (\tfrac{4}{n}-1)I_1]$$
$$\mathrm{II} \equiv (2\partial_i N\bar\Delta u + \partial^j N\bar\nabla_i\partial_j u).\bar\nabla^i u' + NP_{ij}[\bar\nabla^i u'.\bar\nabla^j u'\ +\ 2\bar\nabla^i\partial^j u.\bar\Delta u]$$
$$\mathrm{III} \equiv \partial^h(N\tau)u'.\bar\nabla_h u' + [(\bar\nabla(N\tau) + \mathbf{2}\bar\nabla_i(NK^{ih}))\partial_h u +\ \bar\Delta N u'].\bar\Delta u$$
$$\mathrm{IV} \equiv N^{-1}R^A_{CD\ E}(u)\partial_0 u^C\partial_i u^B\partial_0 u^E\ .\bar\nabla^i u' + R_{CD\ E}\partial_0 u^C\partial^i u^D\partial_i u^E.\bar\Delta u.$$

We note that the terms I and II are quadratic in the second derivatives of u, with coefficient $N\tau$ for I, up to numbers depending only on n. In the case of II the coefficients belong to DN or NP. The term III is bilinear in the first and the second derivatives of u with coefficients $\bar\nabla(NK)$ and $\bar\Delta N$, while IV is linear in second derivatives of u with coefficients cubic in the first derivatives of u and linear in the Riemann tensor of the target metric h.

We note also the following lemma.

Lemma. For an arbitrary map for which the following integrals make sense the following equality holds:

$$\int_{S_t}\bar\Delta u.\bar\Delta u\mu_t = \int_{S_t}|\ \bar\nabla\partial u\ |^2_{g,h}\ \mu_t\ +$$
$$\int_{S_t}\{\bar R^{ij}\partial_i u.\partial_j u \text{ - } R_{A\dot B\ C}\partial^i u^A\partial^j u^B\partial_i u^C.\partial_j u\}\mu_t$$

where $\bar{R}_{ij}$ is the Ricci tensor of the space metric g.

Proof. Stokes' formula gives

$$\int_{S_t} \bar{\nabla}^i \partial_i u . \bar{\nabla}^j \partial_j u \mu_t = -\int_{S_t} \bar{\nabla}^j \bar{\nabla}^i \partial_i u . \partial_j u \mu_t$$

The Ricci formula gives

$$\bar{\nabla}^j \bar{\nabla}^i \partial_i u \equiv \bar{\nabla}^i \bar{\nabla}^j \partial_i u - \bar{R}^{ih} \partial_h u + R_{A\dot{B}\;C} \partial^j u^A \partial^i u^B \partial_i u^C.$$

Another application of Stokes' formula achieves the proof of the lemma.

6 Case of $n \leq 3$

In the case where $n \leq 3$ the Sobolev embedding theorem can be used to estimate the second energy $e^{(1)}(t,u)$ in terms of the H_1 norms of Du and u.

We enunciate and prove a general theorem.

Theorem. (second energy estimate). There exists a number $T > 0$ and a function $C(t)$ continuous in [0,T) such that if **g** satisfies the hypothesis and Riemann(h) is uniformly bounded on the target M then the second energy $y(t)$ satisfies the inequality:

$$y(t) \leq C(t) \quad \text{for} \quad 0 \leq t < T$$

Proof. We first bound the absolute values of the various terms appearing in the right hand side of the energy equality proved in the previous section. We denote generically by C numbers depending only on the dimension n. We denote by $|.|_{\;g,h}$ pointwise norms in the metrics g and h.

We have:

$$|\mathrm{I}| \leq CN \mid \tau \mid (I_0 + I_1).$$

Rather than bounding the absolute value of II we will bound at once its integral. We use the lemma of the previous section which implies

$$\int_{S_t} \mid \bar{\nabla} Du \mid^2_{g,h} \mu_t \leq \int_{S_t} \{ \mid \bar{\Delta} u \mid^{\,2}_{\,g,h} + \mid \mathrm{Ricci(g)}|_g| \, Du \mid^2_{g,h}$$
$$+ |\mathrm{Riemann(h)}|_{\;h}| \, DU \mid^{\,4}_{\,g,h} \} \mu_t.$$

We use the general property of scalar products that $\mid a.b \mid \leq \mid a \mid \mid b \mid \leq \frac{1}{2} \, (\mid a \mid^{\,2} + \mid b \mid^2)$ to obtain

$$\int_{S_t} |\mathrm{II}| \, \mu_t \leq \int_{S_t} \mid II_a \mid \mu_t + \int_{S_t} |\mathrm{II}_b \mid \mu_t$$

with

$$\int_{S_t} |\mathrm{II}_a| \, \mu_t \leq C \int_{S_t} \{ \, [\mid DN \mid_g + N \mid P \mid_g][I_0 + I_1]$$
$$+ \mid DN \mid_{\;g} \mid Ricci(g) \mid_g | \, Du \mid_{g,h} I_0^{1/2} + \mid NP \mid_g | \, Ricci(g) \mid_g | \, Du \mid_{g,h} I_1^{1/2} \; \} \; \mu_t.$$

while

$$\int_{S_t} \mid II_b \mid \mu_t \leq \int_{S_t} \{ [\mid DN \mid_{\;g} + \mid NP \mid_{\;g}] \, |\mathrm{Riemann(h)}|_{\;h}| \, Du \mid^{\,4}_{\,g,h} \} \mu_t$$

The absolute value of III is bounded as follows:

$$|\mathrm{III}| \leq C \mid D(N\tau) \mid_g | \, u' \mid_h I_0^{1/2}] + [(\mid D(N\tau) \mid_g + \mid \bar{\nabla}(NK) \mid_g) \mid Du \mid_{g,h} + \mid \bar{\Delta} N \mid \mid u' \mid_h] I_1^{1/2}.$$

Finally

$$|\mathrm{IV}| \leq CN \mid Riemann(h(u)) \mid_h [| \, Du \mid^2_{g,h} | \, u' \mid_h I_1^{1/2} + \mid Du \mid_{g,h} | \, u' \mid^2_h I_0^{1/2}]$$

The integrals of I, II_a, III can immediately be bounded, for any n, in terms of the first and second energies of u through the use of the Cauchy-Schwarz inequality, if we suppose that DN, NP, their gradients and the Ricci tensor of g are uniformly bounded in g norm on S_t.

We denote by y the second energy, namely we set:

$$y \equiv e^{(1)}(t,u) \equiv y_0 + y_1$$

with

$$y_0(t) \equiv \int_{S_t} I_0 \mu_t, \quad y_1(t) \equiv \int_{S_t} I_1 \mu_t$$

Recall that the first energy was $e(t,u) \equiv e_0 + e_1$ with

$$e_0(t) \equiv \tfrac{1}{2} \int_{S_t} |u'| \, ^2_h \mu_t, \quad e_1(t) \equiv \tfrac{1}{2} \int_{S_t} | \, Du \mid^2_{g,h} \mu_t,$$

We then obtain, omitting to write the explicit dependence on t to abbreviate notations and denoting by C constants depending only on n,

$$\int_{S_t} |\mathrm{I}| \; \mu_t \leq C \; [Sup_{S_t} | \, N\tau \, |][\, y_0 + y_1]$$

$$\int_{S_t} |\mathrm{II}_a| \; \mu_t \leq C\{[\mathrm{Sup}_{S_t} \mid DN \mid_g] \; y_0^{1/2} y_1^{1/2} + [\; \mathrm{Sup}_{S_t} | \, NP \mid_g][y_0 + 2y_1]$$

$$+ \; [\mathrm{Sup}_{S_t} | \, DN \mid_g \mid Ricci(g) \mid_g] e_1^{1/2} y_0^{1/2} + [Sup_{S_t} \mid NP \mid_g | \, Ricci(g) \mid_g] e_1^{1/2} y_1^{1/2} \; .$$

<u>Remark.</u> One can use an L^p norm of Ricci(g) intead of the Sup norm, and estimates of an L^q norms of Du and u'. These norms themselves being estimated in terms of the first and second energies, as we will do later in bounding the integrals of IV and II_b.

We now estimate the integral of III. We find:

$$\int_{S_t} |\mathrm{III}| \; \mu_t \leq C\{[\mathrm{Sup}_{S_t} \mid D(N\tau) \mid] e_0^{1/2} y_0^{1/2}$$

$$+ \; \mathrm{Sup}_{S_t}[| \, D(N\tau \mid_g + \mid \bar{\nabla}(NP) \mid_g] e_1^{1/2} y_1^{1/2} + [\mathrm{Sup}_{S_t} \mid \bar{\Delta} N \mid] e_0^{1/2} y_1^{1/2}.$$

Since IV is cubic in ∂u some further estimates are needed to obtain its bound in terms of e and y. We proceed as follows.

The Cauchy-Schwarz inequality implies:

$$\int_{S_t} |\mathrm{IV}| \; \mu_t \leq \mathrm{C}[\mathrm{Sup}_{S_t} N \mid Riemann(h(u)) \mid_h][\| | \, Du \mid^2_{g,h} | \, u' \mid_h \|_{L^2(g)} \; y_1^{1/2}$$

$$+ \; \| | \, Du \mid_{g,h} | \, u' \mid^2_h \|_{L^2(g)} \; y_0^{1/2}]$$

By Hölder's inequality we have for arbitrary functions F and G on S:

$$\| \, F^2 G \, \|_{L^2(g)} \leq \| \, F^2 \, \|_{L^3(g)} \| \, G \, \|_{L^6(g)}, \quad \text{because} \quad \tfrac{1}{2} = \tfrac{1}{3} + \tfrac{1}{6} \; .$$

This inequality together with $\| \, F^2 \, \|_{L^3} \equiv \| \, F \, \|^2_{L^6}$ gives the estimate

$$\int_{S_t} |\mathrm{IV}| \; \mu_t \leq C \; \mathrm{Sup}_{S_t} \mid Riemann(h(u)) \mid_h \{\| | \, Du \mid_{g,h} \|^2_{L^6(g)} \| \; |u'| \; _h \; \| \; _{L^6(g)} y_1^{1/2}$$

$$+ \; \| | \, Du \mid_{g,h} \|_{L^6(g)} \| | \, u' \mid_h \|^2_{L^6(g)} \; y_0^{1/2}\}.$$

Due to the hypothesis made on the metric g the norms in $L^p(g)$ are equivalent to the norms L^p in the Sobolev regular metric s on S. One can use the Sobolev embedding theorem on (S, e) to estimate the L^6 norm of an arbitrary function F on S in terms of its H_1 norms if $n \leq 3$

$$\| F \|_{L^6} \leq C_s \| F \|_{H_1}, \quad \text{with} \quad \| F \|^2_{H_1} \equiv \| F \|^2_{L^2} + \| DF \|^2_{L^2}$$

where DF is the gradient of the scalar function F. Set

$$F_0 \equiv |u'|_h \equiv \{N^{-2} h_{AB} \partial_0 u^A \partial_0 u^B\}^{1/2}.$$

The gradient of F, a scalar function is independent of the metric of the space, that is $DF \equiv \bar{\nabla} F$. Therefore we can use the Leibnitz rule for covariant derivatives of mappings to obtain:

$$DF_0 = \frac{\bar{\nabla} u'.u'}{|u'|_h} \quad \text{which implies} \quad | DF_0|_g \leq | \bar{\nabla} u' |_{g,h},$$

and, with C_g a number depending only on the equivalence bounds between the metrics g and e,

$$| DF_0 | \leq C_g | \bar{\nabla} u' |_{g,h}, \quad \text{hence} \quad \| DF_0 \|_{L^2} \leq C_g y_0^{1/2}.$$

An analogous reasoning applied to

$$F_1 \equiv | Du |_{g,h}$$

gives

$$\| DF_1 \|_{L^2} \leq C_g y_1^{1/2}.$$

Using these inequalities we obtain a bound in terms of the first energy $e \equiv e(t, u)$ and the second energy $y \equiv e^{(1)}(t, u)$, given by the following estimate (we use the fact that if a and b are positive numbers then $(a+b)^3 \leq 4(a^3+b^3)$)

$$\int_{S_t} |\mathrm{IV}| \, \mu_t \leq C C_g C_h \{e^{3/2} y^{1/2} + y^2\}$$

with

$$C_h \equiv \mathrm{Sup}_{S_t} | \mathit{Riemann}(h(u)) |_h$$

The bound of the integral of $|\mathrm{II}_b|$ is obtained similarly because

$$\int_{S_t} |\mathrm{II}_b| \, \mu_t \leq \int_{S_t} | DN |_g \, C_h \, | Du |^4_{g,h} \, \mu_t$$

$$\leq \mathrm{Sup}_{S_t} [| DN |_g + | NP |_g] C_h \| Du \|_{L^2(g)} \| Du \|^3_{L^6(g)}.$$

Therefore

$$\int_{S_t} |\mathrm{II}_b| \, \mu_t \leq C'_g C_h (e^{1/2} y^{3/2} + e^{3/2} y^{1/2})$$

By the first energy estimate we know that $e \equiv e(t, u)$ is a continuous function of $t \in [0, \infty)$. The obtained inequality give therefore for $y(t)$ a differential inequality of the following type:

$$\frac{dy}{dt} \leq C\{ \alpha y + \beta y^{1/2} + \gamma y^{3/2} + \delta y^2\}$$

The theorem follows from the application of Gromwall's lemma and the fact that the differential equation satisfied by y corresponding to this differential inequality has a continuous solution z on the interval $[0, T)$, for some small enough $T > 0$ which takes the value $z(0) = y(0)$ for $t = 0$.

The expressions for the functions α, β ,γ, δ can be read from the inequalities written above.

The coefficients γ and δ of the nonlinear terms are zero if $C_h = 0$, i.e. if the target is flat. The nonflatness of the target is an obstruction to a global in time estimate.

Remark. The term in $C'_{\mathbf{g}}$ can be expressed differently, using an L^p norm of Ricci(g) intead of the Sup norm, and estimates of an L^q norms of Du and u', estimated again in terms of the first and second energies.

7 Estimate of H_1 Norms

We have seen that the L^2 norms on S_t of DU and $N^{-1}\partial_0 U$ are equal to the energy elements $e_1(t,u)$ and $e_0(t,u)$ respectively. It is not true for the H_1 norms of these quantities compared with the second energy which are defined through covariant mapping derivatives.

For instance we have (cf. Sect. 2)

$$D_i\partial_j U^a \equiv \partial_A i^a D_i \partial_j u^A - K^a_{bc}\partial_i U^b \partial_j U^c.$$

We deduce from this identity and the multiplication properties of Sobolev spaces again an estimate of the H_1 norm of DU on S_t in terms of the first and second energies of u, hence the following lemma.

Lemma. The Cauchy problem for the wave map equation on $S\times\ [\ t_0,\infty)$ has a global solution if its second energy does not blow up in a finite time.

8 Case $n = 1$

In this case the Gagliardo-Nirenberg interpolation inequalities as extended by Aubin 1982 [11] to riemannian manifolds can be used to reduce the degree of the terms in second derivatives appearing in the final estimate. This method was used by Ginibre and Velo (1981) [10] to prove global existence of wave maps on two-dimensional Minkowski space time. However, the interpolation theorem on a compact manifold involves the mean value of the function one wants to estimate, and this poses difficulties. Instead of this interpolation we will use simply the Sobolev embedding theoren of L^3 into W^1_1 when $n = 1$: in this case the Sobolev embedding theorem that there exists a constant C_s, depending only on S and the given metric s, such that

$$\| F \|_{L^3} \leq C_s \| F \|_{W^1_1}, \quad \text{with} \quad \| F \|_{W^1_1} \equiv \| F \|_{L^1} + \| DF \|_{L^1}$$

, where DF is the gradient of the scalar function F. Note that, if a function is in L^6, its square is in L^3. Set

$$F_0 \equiv |u'|^2_h \equiv N^{-2} h_{AB}\partial_0 u^A \partial_0 u^B.$$

The gradient of F, a scalar function is independent of the metric of the space, that is $DF \equiv \bar{\nabla} F$. Therefore we can use the Leibnitz rule for covariant derivatives of mappings to obtain:

$$DF_0 = 2\bar{\nabla}u'.u' \quad \text{which implies} \quad |\ DF_0|_g \leq |\ \bar{\nabla}u'\ |_{g,h}.\ |u'|_h,$$

and, with C_g a number depending only on the equivalence bounds between g and δ,

$$| DF_0 | \leq C_g | \bar{\nabla} u' |_{g,h} . | u' |_h, \quad \text{hence} \quad \| DF_0 \|_{L^1} \leq C_g e_0^{1/2} y_0^{1/2}.$$

An analogous reasoning applied to

$$F_1 \equiv | Du |^2_{g,h}$$

gives

$$\| DF_1 \|_{L^1} \leq C_g e_1^{1/2} y_1^{1/2}.$$

These inequalities lead to a linear inequality for the second energy y on S_t which proves that it does not blow up. The method has been applied to wave maps on Schwarzchild black holes (cf. C-B 1998a [23])

9 Case $n = 2$

The global existence (without uniqueness) of weak solutions of the wave map equation on 2 + 1 dimensional Minkowski space time has been proved by Muller and Struwe 1997 [22] in first energy space. One can hope to prove global existence of strong, unique, solutions using again an interpolation inequality to reduce the differential inequality satisfied by the second energy using the bound of the first.

We recall the general interpolation inequality on a riemannian manifold.

Lemma. (cf. Aubin [11, p 93] or C-B DeWitt [15, p 384]). Let (S, s) be R^n or a compact manifold with or without boundary. Then there exists a constant C_s depending only on (S, s) and n, m, j, q and r such that for all functions $f \in \mathcal{D}(S)$, and satisfying:

$$\bar{f} \equiv \int_S f \mu_s = 0$$

in the case where S is compact without boundary, it holds

$$\| D^j f \|_{L^p} \leq C_s \| D^m f \|^a_{L^r} \| f \|^{1-a}_{L^q}$$

where

$$\tfrac{1}{p} = \tfrac{j}{n} + a\{\tfrac{1}{r} - \tfrac{m}{n}\} + (1-a)\tfrac{1}{q}, \qquad \tfrac{j}{m} \leq a \leq 1, \quad p \geq 1.$$

The inequality is not valid for $a = 1$ if $r = n/(m-j) \neq 1$.

If we suppose that S is R^2 the interpolation theorem can be used with $j = 0$ and $m = 2$ to estimate the L^6 norm of a function $f \in \mathcal{D}$ as follows:

$$\| f \|_{L^6} \leq C_s \| D^2 f \|^{1/3}_{L^2} \| f \|^{2/3}_{L^2}.$$

We can apply such an inequality to the functions $| Du |$ and $|u'|$ to estimate the cube of their L^6 norms in terms of the first energy and linearly in terms of $D^2 | Du |$ and $D^2 | u' |$. If these second derivatives could be estimated with respect to $| D^2 Du |^2$ or $| D^2 u' |$, that is the third energy of u, we could hope to obtain linear differential inequalities for the second and third energies. Unfortunately these estimates are not simple in general. It is possible that estimates of the second and third energies can be obtained though interpolation inequalities for special sources (Minkowski spacetime) and targets (spaces of constant riemannian cuevature). We leave the study to further work.

B. Expanding Universes

We will use refinements of our previous estimates to show that for small data the second energy is bounded in the expanding direction of an expanding universe of dimension $n+1=3$. This energy does not blow up in a finite time if $n+1=4$.

We consider on a manifold S$\times R$ a spacetime metric of the form

$$\mathbf{g} \equiv - N^2 dt^2 + g, \quad \text{with} \quad g \equiv R^2 \sigma$$

The function R depends only on t and is increasing.

The function N and the metric σ satisfy the hypothesis made on N and g in Part A. Moreover the metric σ is uniformly equivalent to the given Sobolev regular metric s for $t \geq t_0$. We also suppose that the upper and lower bounds of N on each S_t are uniformly bounded when t tends to infinity. The behaviour of N depends on the choice of the time parameter t: Ndt is the infinitesimal proper time - cosmic time - on the time line, we choose t to be equivalent to it.

The extrinsic curvature of a submanifold S_t is:

$$K_{ij} \equiv -\frac{1}{2N}\partial_t g_{ij} \equiv -N^{-1}\{R\partial_t R \sigma_{ij} + \frac{R^2}{2}\partial_t \sigma_{ij}\}$$

The mean extrinsic curvature is:

$$\tau \equiv g^{ij} K_{ij} \equiv -N^{-1}\{nR^{-1}\partial_t R + \frac{1}{2}\sigma^{ij}\partial_t \sigma_{ij}\}$$

We suppose that:

$$\partial_t R > 0 \quad \text{for} \quad t > t_0 > 0$$

and we say then that the universe $(S \times R, \mathbf{g})$ is expanding.

1 First Energy Estimate

We have obtained in Part A (corollary of Theorem 1) the equality satisfied by the first energy of a wave map

$$\frac{d}{dt}e(t,u) = \int_{S_t} N\tau\{(\frac{1}{n}-\frac{1}{2})\{\mid Du \mid^2_{g,h} + \frac{1}{2} \mid N^{-1}\partial_0 u \mid^2_h\}\mu_t$$
$$\int_{S_t} N^{-1}\partial_i N \partial^i u . \partial_0 u + NP^{ij}\partial_i u . \partial_j u\}\mu_t .$$

We replace $N\tau$ by its value. Using the fact that R depends only on t we have

$$\frac{d}{dt}e(t,u) = - R^{-1}\partial_t R \int_{S_t}\{(1-\frac{n}{2})\{\mid Du \mid^2_{g,h} + \frac{n}{2} \mid N^{-1}\partial_0 u \mid^2_h\}\mu_t + \mathcal{R}$$

with

$$\mathcal{R} \equiv \int_{S_t} -\frac{1}{2}\varphi\{(\frac{1}{n}-\frac{1}{2}) \mid Du \mid^2_{g,h} + \frac{1}{2} \mid N^{-1}\partial_0 u \mid^2_h\}\mu_t +$$
$$\int_{S_t} N^{-1}\partial_i N \partial^i u . \partial_0 u + NP^{ij}\partial_i u . \partial_j u\}\mu_t .$$

with

$$\varphi \equiv \sigma^{ij}\partial_t \sigma_{ij}.$$

Using the notations e_0 and e_1 of Part A we write the energy equality under the form

$$\frac{d}{dt}(e_0 + e_1) = -R^{-1}\partial_t R\{(2-n)e_1 + ne_0\} + \mathcal{R}.$$

Set

$$f \equiv (e_0 + e_1)R^{2-n},$$

we have the equality:

$$\frac{d}{dt}f = -nR^{-1}\partial_t Re_0 + R^{2-n}\mathcal{R}.$$

Therefore, since we have supposed $\partial_t R \geq 0$:

$$\frac{d}{dt}f \leq R^{2-n}\ \mathcal{R} \leq R^{2-n} \mid \mathcal{R} \mid.$$

We now estimate $\mid \mathcal{R} \mid$.

$$\mid \mathcal{R} \mid \leq Sup_{S_t} \mid \varphi \mid \{ \tfrac{n-2}{2}e_1 + \tfrac{1}{2}e_0\} + Sup_{S_t} \mid DN \mid_g e_0^{1/2}e_1^{1/2} + Sup_{S_t} \mid NP \mid_g e_1.$$

The pointwise g and σ norms of the vector DN and the 2-tensor P are linked by the relations:

$$\mid DN \mid_g = R^{-1} \mid DN \mid_\sigma, \quad \mid P \mid_g = R^{-2} \mid P \mid_\sigma.$$

Since the trace free part P of the extrinsic curvature of S_t in the space time is

$$P_{ij} \equiv R^2 p_{ij}, \quad \text{with} \quad p_{ij} \equiv -\frac{\partial_t \sigma_{ij}}{2N}\ .$$

we have

$$\mid P \mid_g = \mid p \mid_\sigma.$$

There exists therefore a constant C depending only on n such that

$$\frac{df}{dt} \leq CR^{-1}(\Pi + \nu)f$$

with

$$\Pi \equiv R\{\ Sup_{S_t} \mid \varphi \mid + Sup_{S_t} \mid Np \mid_\sigma\}.$$

$$\nu \equiv \mathrm{Sup}_{S_t} \mid DN \mid_\sigma\ .$$

We deduce from the differential inequality satisfied by f and Gromwall's lemma the following bound on f:

$$f(t) \leq f(t_0)\exp\int_{t_0}^t CR^{-1}\ (\Pi + \nu)(\tau)\mathrm{d}\tau.$$

We suppose that Π and ν are uniformly bounded on the interval $[\ t_0, \infty)$ by some constant M, the bound on f is then

$$f(t) \leq f(t_0)\exp 2CM\int_{t_0}^t\ R^{-1}(\tau)\mathrm{d}\tau.$$

We have proved the following theorem.

Theorem. If on the interval $[t_0, \infty)$ the functions Π and ν are uniformly bounded and R^{-1} is integrable then the product $f \equiv R^{2-n}(e_0+e_1)$ is uniformly bounded on this interval. Setting

$$\int_{t_0}^\infty R^{-1}(t)dt = \rho$$

it holds

$$e(t,u) \leq R(t)^{n-2}\ell \exp(2CM\rho), \quad \text{with} \quad \ell \equiv e(0,u)R(0)^{2-n}$$

Remark 1. If $\mathbf{g}$ is a Robertson-Walker metric, i.e. if $N = 1$ and the metric σ is independent of t then the gradient of N and the trace free part P of the extrinsic curvature are identically zero as well as the full extrinsic curvature of (S,σ) in the spacetime metric $-dt^2 + \sigma$. We have then $M \equiv 0$.

Remark 2.We introduce the following notations

$$E_0 \equiv \tfrac{1}{2} \parallel u' \parallel^2_{L^2}\ , \quad u' \equiv N^{-1}\partial_t u, \quad E_1 \equiv \tfrac{1}{2} \parallel Du \parallel^2_{L^2},$$

where the pointwise norms are taken in the metrics σ and h and the volume element in the metric σ. Using the identities

$$\mid Du \mid_{g,h} = R^{-1} \mid Du \mid_{\sigma,h}, \quad \mu_t = R^n \mu_\sigma$$

we find

$$E_1 \equiv R^{2-n} e_1, \quad E_0 = R^{-n} e_0.$$

We see that if $f \equiv R^{2-n}(e_0 + e_1)$ is uniformly bounded then E_1 and $R^2 E_0$ are uniformly bounded, i.e. E_0 decays like R^{-2}.

2 Second Energy Estimate

Using the same arguments than in the previous section we find that the second energy $e^{(1)}(t,u) \equiv y \equiv y_0 + y_1$ satisfies the equality:

$$\frac{dy}{dt} = \text{-}\ R^{-1}\partial_t R\{(n+2)y_0 + (4-n)y_1\} + \mathcal{S}$$

with

$$\mathcal{S} \equiv \int_{S_t} \{\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}\}\ \mu_t,$$

where the notations are those of Part A except for I which reduces now to

$$\mathrm{I} \equiv \text{-}\ \tfrac{1}{2}\varphi\ [(1 + \tfrac{\mathbf{2}}{n})\ I_0 + (\tfrac{4}{n} - 1)I_1]$$

We set:

$$z \equiv R^{4-n} y$$

then z satisfies the equality

$$\frac{dz}{dt} = R^{4-n}\{-R^{-1}\partial_t R(2n-2)y_0 + \mathcal{S}\}.$$

and therefore, in an expanding universe, the inequality

$$\frac{dz}{dt} \leq R^{4-n}\mathcal{S}.$$

We now bound the various terms of $R^{4-n}\mathcal{S}$. We find

$$R^{4-n} \int_{S_t} \mid \mathrm{I} \mid\ \mu_t \leq C\ \mathrm{Sup}_{S_t} \mid \varphi \mid z.$$

To write the following bounds we use the relation between the g and σ norms of vectors and tensors noted before.

We also note that the covariant derivatives in the metrics g and σ are the same and so

$$Ricci(g) = Ricci(\sigma), \quad \text{hence} \quad \mid Ricci(g) \mid_g = R^{-2} \mid Ricci(\sigma) \mid_\sigma$$

$$\bar{\Delta} N \equiv R^{-2}\Delta_\sigma N$$

We find:

$$R^{4-n}\int_{S_t}|\mathrm{II}_a|\ \mu_t \leq C\ [\ R^{-1}\ \mathrm{Sup}_{S_t}\mid DN\mid_\sigma + [\mathrm{Sup}_{S_t}|\ Np\mid_\sigma]z$$

$$+\ [\ R^{-1}\mathrm{Sup}_{S_t}(|\ DN\mid_{\ \sigma}\mid Ricci(\sigma)\mid_\sigma)$$

$$+\mathrm{Sup}_{S_t}\mid Np\mid_\sigma|\ Ricci(\sigma)\mid_\sigma]R^{-n/2}e_1^{1/2}z^{1/2}.$$

We now use the fact that $D(N\tau)=D\varphi$ to obtain the bound

$$\int_{S_t}|\mathrm{III}|\ \mu_t \leq C\{[\mathrm{Sup}_{S_t}\mid D\varphi\mid_\sigma]R^{1-(n/2)}e_0^{1/2}z^{1/2}$$

$$+\ \mathrm{Sup}_{S_t}[|\ D(N\tau\mid_\sigma + \mid\bar\nabla(Np)\mid_\sigma]R^{1-(n/2)}e_1^{1/2}z^{1/2}$$

$$+\ [\mathrm{Sup}_{S_t}\mid \Delta_\sigma N\mid]R^{-n/2}e_0^{1/2}z^{1/2}.$$

We now search for a bound of the integral of IV, using Sobolev estimates relative to the fixed metric s which is uniformly equivalent to σ, but not to g. We had:

$$\int_{S_t}|\mathrm{IV}|\ \mu_t \leq C\ C_h\{\||\ Du\mid_{g,h}\|^2_{L^6(g)}\|\ |u'|_{\ h}\ \|_{\ L^6(g)}y_1^{1/2}$$

$$+\ \||\ Du\mid_{g,h}\|_{L^6(g)}\||\ u'\mid_h\|^2_{L^6(g)}\ y_0^{1/2}\}.$$

with

$$C_h\equiv \mathrm{Sup}_{S_t}\mid Riemann(h(u))\mid_h$$

Denoting by $|.|$ a pointwise norm in σ and h and $\|.\|$ a norm in the volume element of σ we have

$$|\ Du\mid_{\ g,h}=R^{-1}\mid Du\mid, \qquad \||\ Du\mid_{\ g,h}\|_{\ L^p(g)}= R^{-1+(n/p)}\ \||\ Du\ |\|_{L^p}.$$

$$|\ D^2u\mid_{g,h}= R^{-2}\mid D^2u\mid, \qquad \||\ D^2u\mid_{g,h}\|_{L^p(g)}= R^{-2+(n/p)}\ \||\ D^2u\ |\|_{L^p}.$$

We recall the estimate resulting from the Sobolev embedding theorem, the uniform equivalence of the metrics g and s and the fact that the norm of the gradient is less than or equal to the gradient of the norm:

$$\||\ Du\ |\|_{\ L^6}\leq C_gC_s\{\||\ Du\ |\|_{L^2}+\||\ D^2u\ |\|_{L^2}$$

It implies

$$R^{1-(n/6)}\ \||\ Du\mid_{\ g,h}\|_{L^6(g)}\leq$$

$$\leq C_gC_s\{R^{1-(n/2)}\ \||\ Du\mid_{g,h}\|_{L^2(g)} + R^{2-(n/2)}\ \||\ D^2u\mid_{g,h}\|_{L^2(g)}\},$$

that is,

$$\||\ Du\mid_{\ g,h}\|_{L^6(g)}\leq C_gC_s\{R^{-(n/3)}\ \||\ Du\mid_{g,h}\|_{L^2(g)} + R^{1-(n/3)}\ \||\ D^2u\mid_{g,h}\|_{L^2(g)}\}.$$

The Sobolev inequality applied to $|u'|$ together with

$$|u'|_{\ h}=|\ u'\ |, \qquad \||u'|_{\ g,h}\|_{\ L^p(g)}= R^{n/p}\ \||\ u'\ |\|_{L^p}.$$

$$|\ Du'\mid_{g,h}= R^{-1}\mid Du'\mid, \qquad \||\ Du'\ |\|_{L^p}= R^{-1+(n/p)}\ \||\ D^2u\ |\|_{L^p}$$

gives the same inequality for L^6 norms when Du is replaced by u'.

We obtain the following bound for the integral of IV by using the definition of the first energy e. We denote by $C_{\mathbf{g}}$ constants depending only on n, s and the uniform bounds of g and N:

$$\int_{S_t}|\mathrm{IV}|\ \mu_t \leq C_{\mathbf{g}}\ C_h\{R^{-n}e^{3/2}+R^{3-n}y^{3/2}\}y^{1/2}.$$

We now use the definitions of $f \equiv R^{2-n}e$ and $z \equiv R^{4-n}y$, together with the elementary fact holding for any pair of positive numbers $(a+b)^3 \leq 4$ (a^3+b^3) to write

$$R^{4-n} \int_{S_t} |\mathrm{IV}| \ \mu_t \leq CC_g C_\delta \ C_h R^{-1}\{f^{3/2}z^{1/2} + z^2\}$$

The previously used inequalities give also

$$R^{4-n} \int_{S_t} |\mathrm{II}_b \mid \mu_t \leq \mathrm{Sup}_{S_t}[R^{-1} \mid DN \mid_\sigma + \mid Np \mid_\sigma] C_h\{ f^{3/2}z^{1/2} + f^{1/2}z^{3/2}\}.$$

Using again the definition of f and assembling previous results we obtain the following theorem.

Theorem. When $n = 2$ or 3 the second energy $e^{(1)}(t,u)$ is such that its product z by R^{4-n} satisfies the differential inequality

$$\frac{dz}{dt} \leq C\{\alpha z^{1/2} + \beta z + \gamma z^{3/2} + \delta z^2\},$$

where α, β, γ, δ are the following positive functions of t

$$\alpha \equiv \{ \ R^{-1}[\ R^{-1} Sup_{S_t}(\mid DN \mid_\sigma \mid Ricci(\sigma) \mid_\sigma) + \mathrm{Sup}_{S_t}(\mid Np \mid_\sigma \mid Ricci(\sigma) \mid_\sigma)]$$
$$+ \ \mathrm{Sup}_{S_t}[\mid D\varphi \mid_\sigma + \mid \bar{\nabla}(Np) \mid_\sigma] + R^{-1}[Sup_{S_t} \mid \Delta_\sigma N \mid]\} f^{1/2} + C_{\mathbf{g}} C_h R^{-1} f^{3/2}$$

$$\beta \equiv R^{-1} \ \mathrm{Sup}_{S_t} \mid DN \mid_\sigma + \ Sup_{S_t}[\mid Np \mid_\sigma + \mid \varphi \mid]$$
$$+ \ \mathrm{Sup}_{S_t}[R^{-1} \mid DN \mid_\sigma + \mid Np \mid_\sigma] C_h f^{3/2},$$

$$\gamma \equiv \mathrm{Sup}_{S_t}[R^{-1} \mid DN \mid_\sigma + \mid Np \mid_\sigma] C_h \ f^{1/2}$$

$$\delta \equiv C_{\mathbf{g}} C_h \ R^{-1}.$$

3 Global Existence Theorem

It follows from the general theory that the wave map equation has global solutions on $S \times R$, with S of dimension $n \leq 3$, if its first and second energies e and $e^{(1)}$ do not blow up in finite time. We have given conditions under which the function $f \equiv R^{2-n}e$ remains finite when t tends to infinity.

It is possible to study the behaviour of $e^{(1)}$ by using the differential inequality satisfied by $z \equiv R^{4-n}e^{(1)}$ and the previously found estimate of the first energy as we have done (C-B 1998c [24]) in the case of a Robertson-Walker spacetime. However, a better handling of the conditions to impose on the initial data is obtained by estimating directly the sum of the weighted energies, namely

$$w \equiv f + z.$$

We make the following hypothesis.

Hypothesis

$\mathcal{H}_1$. The space time metric $\mathbf{g}$ on $S \times I$, with I the interval $[\ t_0, \infty)$ and S a smooth manifold of dimension 2 or 3 is

$$\mathbf{g} = - \ N^2 dt^2 + R^2 \sigma.$$

The lapse N is uniformly equivalent to a positive constant. Its space gradient DN is uniformly bounded in σ norm. Its Laplacian in the metric σ

is uniformly bounded. The product by R of the Ricci tensor of σ is uniformly bounded in σ norm.

The product by R^{-1} of the extrinsic curvature of (S, σ) as submanifold of $(S \times I, -N^2 dt^2 + \sigma)$ is uniformly bounded in σ norm as well as its first space derivative.

$\mathcal{H}_2$. The properly riemannian metric σ is uniformly equivalent to a given riemannian metric s on S, Sobolev regular.

$\mathcal{H}_3$. The coefficient R is a C^1 positive and nondecreasing function depending only upon t. The function R^{-1} is integrable on I. We set

$$\int_{t_0}^{\infty} R^{-1}(t)dt = \rho.$$

$\mathcal{H}_4$. The target (M, h) is a smooth complete riemannian manifold with curvature uniformly bounded in h norm.

We write a differential inequality for $w \equiv f + z$ using the first energy inequality

$$\frac{df}{dt} \leq CR^{-1}(\Pi + \nu)f$$

and the second energy inequality obtained in the previous section together with the obvious fact that $f \leq w$ and $z \leq w$. We obtain an inequality of the type

$$\frac{dw}{dt} \leq R^{-1}\{Aw + K_h w^2\},$$

where the constant A depends only on the uniform bounds specified in the hypothesis $\mathcal{H}_1$ that we made on N and the metric σ. The constant K_h is a product by C_h, the bound of the h norm of the Riemann tensor of h, with a constant depending only on the uniform bounds of the hypothesis $\mathcal{H}_1$ and $\mathcal{H}_2$ on N, σ and δ.

It results from Gromwall's lemma that

$$w(t) \leq W(t),$$

where W is the solution which takes for t$= t_0$ the value

$$W(t_0) \equiv W_0 = w(t_0) \equiv f(t_0) + z(t_0)$$

of the following differential equation:

$$\frac{dW}{dt} = R^{-1}\{AZ + K_h Z^2\}$$

$$F(W) \equiv \int_{W_0}^{W} \frac{d\xi}{A\xi + K_h \xi^2} = \int_{t_0}^{t} \frac{d\tau}{R(\tau)} \equiv \rho(t).$$

This implicit equation for W as a function of $\rho(t)$, i.e. of t, has a bounded solution for all $t \in I \equiv [t_0, \infty)$ if the curve $Y = F(W)$ drawn in the cartesian plane is cut by the lines $Y = \rho(t)$, $t \in I$. It is elementary to check that the function $Y = F(W)$ is monotonously increasing and convex towards positive Ys. When W tends to infinity, $F(W)$ tends to infinity if $K_h = 0$, i.e. if the target is flat: in this case the line $Y = \rho(t)$ cuts the curve $Y = F(W)$, in one point $W(t)$ and therefore the solution $W(t)$ of the implicit equation for W exists, bounded, for every $t \in I$. In the case where $K_h \neq 0$ and $W_0 \neq 0$,

$F(W)$ tends to a finite limit $\mu(W_0)$ when W tends to infinity. The implicit equation for W has a bounded solution for all $t \in I$ if ρ is less than this limit.

Lemma. The limit $\mu(W_0)$ is given by

$$\mu(W_0) \equiv \int_{W_0}^{\infty} \frac{d\xi}{A\xi + K_h\xi^2} = A^{-1} log\{\frac{AK_h^{-1}W_0}{1+AK_h^{-1}W_0}\}, \quad \text{if A} \neq 0,$$

$$\mu(W_0) = \frac{1}{K_h W_0} \qquad \text{if A} = 0.$$

Theorem. Let the spacetime $(S \times I, \mathbf{g})$, $I \equiv [t_0, \infty)$ and the target (M, h) satisfy the hypothesis $\mathcal{H}$. There is a neighbourhood ω of zero in $H_1 \times H_1$ such that for all wave maps initial data (φ,ψ) with ($D\Phi$,Ψ) $\in \omega$ the wave map equation has a global solution on $S \times I$.

Proof. We see on the value of $\mu(W_0)$ that for any given number ρ there is a number $\ell > 0$ such that $W_0 < \ell$ implies $\mu(W_0) > \rho$. For wave maps initial values such that

$$w(t_0) \equiv f(t_0) + z(t_0) \equiv R(t_0)^{2-n} e(t_0) + R(t_0)^{4-n} e^{(1)}(t_0) < \ell.$$

We have seen in Sect. 6 of Part A that the above inequality is equivalent to a bound of the H_1 norms of $D\Phi$ and Ψ.

Acknowledgements. Discussions and suggestions by V. Moncrief and L. Andersson are gratefully aknowledged.

Part of this work was completed in the stimulating atmosphere of the program "Classical and Quantum strong gravitational fields" organised by A. Ashtekar and J. Isenberg at the Institute for Theoretical Physics of the University of California, Santa Barbara.

References

1. J. Leray Hyperbolic differential equations I.A.S. Princeton, 1952.
2. P. Dionne Sur les problèmes hyperboliques bien posés J. Anal. Math. Jerusalem 1 1962 1-90.
3. Y. Choquet-Bruhat Partial Differential Equations on a Manifold, Batelle Rencontres 1967, C. DeWitt and J. Wheeler ed. Benjamin.
4. R. Geroch Domains of dependence J. Math. Phys. **11**1970 437-449.
5. Y. Choquet-Bruhat C^∞ solutions of non linear hyperbolic equations Gen. Rel. and Gravitation **2** 1972 359-362.
6. Y. Nutku Harmonic maps in physics, Ann. Inst. Poincaré, **A 21**, 1974 175-183.
7. C. W. Misner Harmonic maps as models of physical theories, Phys. Rev. D. 1978, 4510-4524.
8. Y. Choquet-Bruhat, D. Christodoulou and M. Francaviglia Cauchy data on a manifold, Ann. Inst. Poincaré X**XXI** n°4, 1979 399-414.
9. Gu Chaohao On the Cauchy problem for harmonic maps on two dimensional Minkowski space, Comm. Pure and App. Maths **33**, 1980 727-737.
10. J. Ginibre and G. Velo The Cauchy problem for the O(N), CP(N-1) and GC(N,p) models, Ann. of Phys. **142**, n°2, 1982 393-415.
11. T. Aubin Nonlinear analysis and Monge Ampere equations, Springer 1982.

12. V. Moncrief Reduction of Einstein's equations for vacuum spacetimes with spacelike U(1) isometry groups, Ann. of Phys. **167** n°1, 1986 118-142.
13. Y. Choquet-Bruhat Hyperbolic harmonic maps, Ann Inst. Poincaré **46** n°1 1987, 97-111.
14. J. Shatah Weak solutions and development of singularities in the SU(2) σ-model, Comm. Pure App. Math, **41** 1988 459-469.
15. Y. Choquet-Bruhat and C. DeWitt-Morette Analysis, Manifolds and Physics II North Holland 1989.
16. Y. Choquet-Bruhat and Gu Chaohao Global Existence of Harmonic Maps on Minkowski Spacetime M_3, C.R. Acad. Sci. Paris **308**, Serie I 1989, 167-170.
17. D. Christodoulou and A. Talvilar-Zadeh On the regularity of spherically symmetric wave maps, Comm. Pure Ap. Math. **46** , 1993, 1041-1091.
18. S. Klainerman and M. Machedon Spacetime estimates for null forms and the local existence theorem, Com. Pure App. Math. **46,** 1993, 1221-1268.
19. C. Sogge On local existence for wave equations satisfying variable coefficients null conditions, Comm. Part. Diff. Equ. **18** 1993, 1795-1821.
20. Y. Choquet-Bruhat and V. Moncrief Existence theorem for Einstein's equations with one parameter isometry group, in Proc. Symp. Amer. Math. Soc. **59**, Brezis and Segal ed. 1996, 61-80.
21. J. Shatah and A. Tahvildar-Zadeh Non uniqueness and development of singularities for the harmonic maps of the Minkowski space, preprint.
22. S. Muller and M. Struwe, Global existence of wave maps in 2+1 dimensions with finite energy data, Topological Methods in Nonlinear Analysis **7** n°2, 1996, 245-261.
23. Y. Choquet-Bruhat Wave Maps in General Relativity, in "On Einstein path", A. Harvey ed. Springer 1998a, 161-185.
24. Y. Choquet-Bruhat Global existence of wave maps, Rendi Conti Circ. Mat. Palermo Serie II suppl **57** 1998c, 143-152.

Einstein's Equations and Equivalent Hyperbolic Dynamical Systems

Arlen Anderson[1], Yvonne Choquet-Bruhat[2], and James W. York Jr.[1]

[1] Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC 27599-3255, USA
[2] Gravitation et Cosmologie Relativiste, t.22-12, Un. Paris VI, Paris 75252 France

Abstract. We discuss several explicitly causal hyperbolic formulations of Einstein's dynamical $3+1$ equations in a coherent way, emphasizing throughout the fundamental role of the "slicing function," α—the quantity that relates the lapse N to the determinant of the spatial metric $\bar{g}$ through $N = \bar{g}^{1/2}\alpha$. The slicing function allows us to demonstrate explicitly that every foliation of spacetime by spatial time-slices can be used in conjunction with the causal hyperbolic forms of the dynamical Einstein equations. Specifically, the slicing function plays an essential role (1) in a clearer form of the canonical action principle and Hamiltonian dynamics for gravity and leads to a recasting (2) of the Bianchi identities $\nabla_\beta G^\beta{}_\alpha \equiv 0$ as a well-posed system for the evolution of the gravitational constraints in vacuum, and also (3) of $\nabla_\beta T^\beta{}_\alpha \equiv 0$ as a well-posed system for evolution of the energy and momentum components of the stress tensor in the presence of matter, (4) in an explicit rendering of four hyperbolic formulations of Einstein's equations with only physical characteristics, and (5) in providing guidance to a new "conformal thin sandwich" form of the initial value constraints.

1 Introduction

Einstein's equations have, for much of the history of general relativity, been explored very fruitfully in terms of their concise and elegant statements characterizing the geometry of four dimensional pseudo-riemannian geometries. Such geometries depict possible physical spacetimes containing only "the gravitational field itself." The variety and properties of these "empty" spacetimes is truly astonishing. Quasi-local geometric entities such as trapped surfaces and event horizons have become familiar. It is now firmly established that large-scale topological and geometrical features of spacetime are, indeed, subjects of physical inquiry. The nature and distribution of "matter" at stellar scales and upward has also brought particle physics and hydrodynamics to the fore.

During these years, however, a steady development of the "space-plus-time" or $3+1$ view of spacetime geometry has also occured. Here one views general relativity as "geometrodynamics" in the parlance of John Wheeler [1]. The emphasis, in the canonical or Hamiltonian explication of geometrodynamics given by Arnowitt, Deser, and Misner ("ADM") [2] and by Dirac [3,4], is on the evolving intrinsic and extrinsic geometry of spacelike hypersurfaces

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 30–54, 2000.

which determine, by knowledge of the appropriate initial data and by classical causality, the spacetime "ahead" (and "behind"), if spacetime is globally hyperbolic, an assumption we adopt throughout.

Underlying and preceding geometrodynamics and Hamiltonian methods, however, was the basic realization that four of the ten Einstein vacuum equations are nonlinear constraints on the initial Cauchy data, which play such a decisive role in defining the later canonical formalism [5]. The Cauchy problem, constraints plus evolution, was shown to be well-posed in the modern sense of nonlinear partial differential equations [6,7,8,9]. This train of progress was marked by early work of Darmois and Lichnerowicz [8], and brought to definitive development by one of us [6,7].

At this writing, with accurate three-dimensional simulations using the full Einstein equations, with and without the presence of stress-energy sources, becoming essential for realistic studies of gravity waves, high energy astrophysics, and early cosmology, studies of Einstein's equations of evolution in $3+1$ form have blossomed [11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26, 27,28] (an incomplete sample — see also [10]). Hyperbolic forms, especially first-order symmetrizable forms possessing only physically causal directions of propagation, have undergone very significant development in the past few years.

Specifically, in this chapter we describe in detail several explicitly causal hyperbolic formulations of Einstein's dynamical $3+1$ equations by following a path that can be viewed as lighted by the "slicing function," α — the quantity that relates the lapse N to the determinant of the spatial metric $\bar{g}$ through $N = \bar{g}^{1/2}\alpha$. This representation of the lapse function was presented in [18]. The slicing function allows us to demonstrate explicitly that no foliation of spacetime by spatial time-slices can be an obstacle to the causal hyperbolic forms of the dynamical Einstein equations. The slicing function plays an essential role (1) in a more precise form of the canonical action principle and canonical dynamics for gravity, (2) leads to a recasting of the Bianchi identities $\nabla_\beta G^\beta{}_\alpha \equiv 0$ as a well-posed system for the evolution of the gravitational constraints in vacuum and also (3) of $\nabla_\beta T^\beta{}_\alpha \equiv 0$ as a well-posed system for evolution of the energy and momentum components of the stress tensor in the presence of matter, (4) in an explicit display of four hyperbolic formulations of Einstein's equations with only physical characteristics, and (5) even in providing guidance to a new elliptic "conformal thin sandwich" form of the initial value constraints.

We recall that the proof of the existence of a causal evolution in local Sobolev spaces of $\bar{g}$ and its extrinsic curvature (second fundamental tensor) K into an Einsteinian spacetime does not result directly from the equations giving the time derivatives of $\bar{g}$ and K in terms of space derivatives of these quantities in a straightforward $3+1$ decomposition of the Ricci tensor of the spacetime metric, which contains also the lapse and shift characterizing the time lines. These equations do not appear as a hyperbolic system for arbitrary lapse and shift, in spite of the fact that their characteristics are only the light cone and the time axis [25].

We now turn to notational matters, conventions, and to the $3+1$ decomposition of the Riemann and Ricci tensors. We assume here and throughout the sequel that the spacetime $V = M \times \mathbb{R}$ is endowed with a metric g of signature $(-,+,+,+)$ and that the time slices are spacelike, that is, have signature $(+,+,+)$. These assumptions are not restrictive for globally hyperbolic (pseudo-riemannian) spacetimes.

We choose on V a moving coframe such that the dual vector frame has a time axis orthogonal to the slices M_t while the space axes are tangent to them. Specifically, we set

$$\begin{aligned} \theta^0 &= dt \, , \\ \theta^i &= dx^i + \beta^i dt \, , \end{aligned} \tag{1}$$

with $t \in \mathbb{R}$ and x^i, $i = 1,2,3$ local coordinates on M. The Pfaff or convective derivatives ∂_α with respect to θ^α are

$$\begin{aligned} \partial_0 &\equiv \frac{\partial}{\partial t} - \beta^i \partial_i \\ \partial_i &\equiv \frac{\partial}{\partial x^i} \end{aligned} \tag{2}$$

In this coframe, the metric g reads

$$ds^2 = g_{\alpha\beta}\theta^\alpha\theta^\beta \equiv -N^2(\theta^0)^2 + g_{ij}\theta^i\theta^j \; . \tag{3}$$

The t-dependent scalar N and space vector β are called the lapse function and shift vector of the slicing. These quantities were explicitly identified in [7] and play prominent roles in all subsequent $3+1$ formulations. Any spacetime tensor decomposes into sets of time dependent space tensors by projections on the tangent space or the normal to M_t.

We define for any t-dependent space tensor T another such tensor of the same type, $\bar{\partial}_0 T$, by setting

$$\bar{\partial}_0 \equiv \frac{\partial}{\partial t} - \pounds_\beta \, , \tag{4}$$

where $\pounds_\beta$ is the Lie derivative on M_t with respect to β.

Notice that in our foliation-adapted basis (1) and (2), that if g denotes the spacetime metric and $\bar{g}$ the space metric, then we have ($g_{0i} = g^{0i} = 0$ in our frames):

$$g_{ij} = \bar{g}_{ij} \; ; \quad g^{ij} = \bar{g}^{ij} \; . \tag{5}$$

(Greek indices range $\{0,1,2,3\}$ while latin ones are purely spatial.) Hence, *no overbars will be used to denote components of the spatial metric.* On the other hand, for the determinants, we have $(-\det g) = N^2(\det \bar{g})$. Therefore we shall use overbars on spatial metric determinants; for example $\bar{g}^{1/2} \equiv (\det \bar{g})^{1/2}$.

Likewise, to distinguish the purely spatial components of the spacetime Ricci tensor (say), we shall write $R_{ij}(g)$, while for the space Ricci tensor we shall write $R_{ij}(\bar{g})$. In general, of course, $R_{ij}(g) \neq R_{ij}(\bar{g})$. The Levi-Civita connection of g is denoted by ∇ and that of $\bar{g}$ by $\bar{\nabla}$.

With the convention

$$\nabla_\alpha \sigma_\beta \equiv \partial_\alpha \sigma_\beta - \sigma_\rho \gamma^\rho{}_{\beta\alpha} \ , \tag{6}$$

and the definitions

$$\gamma^\alpha{}_{\beta\gamma} = \Gamma^\alpha{}_{\beta\gamma} + g^{\alpha\delta} C^\varepsilon{}_{\delta(\beta} g_{\gamma)\varepsilon} - \frac{1}{2} C^\alpha{}_{\beta\gamma} \ , \tag{7}$$

$$d\theta^\alpha = -\frac{1}{2} C^\alpha{}_{\beta\gamma} \theta^\beta \wedge \theta^\gamma \ , \tag{8}$$

we have for the connection coefficients (Γ denotes an ordinary Christoffel symbol)

$$\gamma^i{}_{jk} = \Gamma^i{}_{jk}(g) = \Gamma^i{}_{jk}(\bar{g}) \tag{9}$$

$$\gamma^i{}_{0k} = -NK^i{}_k \ , \quad \gamma^i{}_{j0} = -NK^i{}_k + \partial_j \beta^i \ , \quad \gamma^0{}_{ij} = -N^{-1} K_{ij} \tag{10}$$

$$\gamma^i{}_{00} = N\partial^i N \ , \quad \gamma^0{}_{0i} = \gamma^0{}_{i0} = \partial_i \log N \ , \quad \gamma^0{}_{00} = \partial_0 \log N \ . \tag{11}$$

Observe that if α is a space scalar of weight -1, we have

$$\bar{\nabla}_i \alpha = \partial_i \alpha + \alpha \Gamma^k{}_{ki}(g) \ = \ \partial_i \alpha + \alpha \partial_i \log \bar{g}^{1/2} \ , \tag{12}$$

$$\pounds_\beta \alpha = \beta^i \bar{\nabla}_i \alpha + \alpha \bar{\nabla}_i \beta^i \ . \tag{13}$$

The Riemann tensor is fixed by

$$\left(\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha\right) V^\gamma = V^\delta R^\gamma{}_{\delta\alpha\beta} \tag{14}$$

while the Ricci tensor is $R_{\delta\beta} \equiv R^\gamma{}_{\delta\gamma\beta}$.

The $3+1$ decompositions of the Riemann and Ricci tensors are

$$R_{ijkl}(g) = R_{ijkl}(\bar{g}) + 2K_{i[k} K_{l]j} \ , \tag{15}$$

$$R_{0ijk}(g) = 2N \bar{\nabla}_{[j} K_{k]i} \ , \tag{16}$$

$$R_{0i0j}(g) = N\left(\bar{\partial}_0 K_{ij} + NK_{ik} K^k{}_j + \bar{\nabla}_i \partial_j N\right) \ . \tag{17}$$

One can then obtain for the Ricci tensor

$$R_{ij}(g) = R_{ij}(\bar{g}) - N^{-1} \bar{\partial}_0 K_{ij} + KK_{ij} - 2K_{ik} K^k{}_j - N^{-1} \bar{\nabla}_i \partial_j N \ , \tag{18}$$

$$R_{0j}(g) = N\left(\partial_j K - \bar{\nabla}_h K^h{}_j\right) \ , \tag{19}$$

$$R_{00}(g) = N\left(\partial_0 K - NK_{ij} K^{ij} + \triangle_{\bar{g}} N\right) \ , \tag{20}$$

where $K \equiv K^i{}_i$ and $\triangle_{\bar{g}} \equiv g^{ij} \bar{\nabla}_i \bar{\nabla}_j$. Finally, we note

$$G^0{}_0 = \frac{1}{2}\left(K_{ij} K^{ij} - K^2 - R(\bar{g})\right) \ . \tag{21}$$

2 Every Time Slicing Is "Harmonic"

The standard statement of the harmonic time-slicing condition is, that on a $t = \text{const.}$ time slice, $\bar{\partial}_0[(-g)^{1/2}g^{00}] = 0$. (This is equivalent, in a coordinate basis, to $\partial_\mu[(-g)^{1/2}g^{\mu t}] = 0$.) Friedrich observed in [24] that the right hand side of these equations could be a given function of (t, x^i)[1]. (See also [19].) Therefore, the standard harmonic condition expressed in $3+1$ form, $\bar{\partial}_0 N + N^2 K = 0$, can be written as a generalized "harmonic" condition

$$\bar{\partial}_0 N + N^2 K = Nf \ , \tag{22}$$

where $f(t,x)$ is a known function. Specifically, introduce $\alpha(x,t)$ such that $\bar{\partial}_0 \log \alpha = f$, then (22) becomes

$$\bar{\partial}_0 N + N^2 K = N \bar{\partial}_0 \log \alpha \ , \tag{23}$$

from which the identity

$$\bar{\partial}_0 \log \bar{g}^{1/2} = -NK \ , \tag{24}$$

allows us to see that

$$N = \bar{g}^{1/2} \alpha \ . \tag{25}$$

We shall call $\alpha(x,t)$ the "slicing function;" it is a freely given scalar density of weight -1.

It is clear that any $N > 0$ on a given time slice $t = t_0$ can be written in the form $N_{t_0} = \bar{g}_{t_0}^{1/2} \alpha(t_0, x)$ for some $\alpha > 0$ provided that $g_{ij}(t_0, x)$ is a proper riemannian metric. Introducing "harmonic" time-slicing is thus a simple matter. It is not, however, known at present how to construct a specific long-time foliation from general rules telling how to specify $\alpha(t,x)$. However, many foliations can be constructed in a "step-by step" fashion (numerical time steps) provided certain obvious conditions are met. For example, an elliptic condition on N can determine $\alpha(t)$ on a sequence of time slices if the condition does not couple to variables that disturb the characteristic directions of the hyperbolic equations. (The same is true for the shift vector β^i.) Alternatively, we can try educated guesses for $\alpha(t,x)$.

As it stands, (23) is clearly a speed zero (with respect to ∂_0) hyperbolic equation. However, this equation and Einstein's equations lead to a second order equation in space and time that propagates N along the light cone. This result brings into sharp relief the congruence of (23) with the propagation on the light cone of other variables.

The trace of $R_{ij}(g)$ gives an equation for $\bar{\partial}_0 K$ [19]

$$\bar{\partial}_0 K = -\triangle_{\bar{g}} N + \left[R(\bar{g}) + K^2 - R^k{}_k(g) \right] N \ , \tag{26}$$

[1] Just as in electrodynamics, $\nabla^\mu A_\mu = 0 \rightarrow \nabla^\mu A_\mu{}' = \ell(t,x) \neq 0$ is perfectly acceptable as a "Lorentz gauge" if ℓ is known.

where $\triangle_{\bar{g}}$ in (26) denotes the Laplacian $g^{ij}\bar{\nabla}_i\bar{\nabla}_j$. Taking the time derivative of (23) and eliminating $\bar{\partial}_0 K$ with (26) shows that N obeys the non-linear wave equation

$$\bar{\Box}_g N + R^k{}_k(g)N - R(\bar{g})N - N\bar{\partial}_0 \log\alpha + (\bar{\partial}_0^2 \log\alpha)N^{-1} = 0 \,, \qquad (27)$$

where we wrote our wave operator or "d'Alembertian" as $\bar{\Box}_g = -(N^{-1}\bar{\partial}_0)^2 + \triangle_{\bar{g}}$. The characteristic cone of $\bar{\Box}_g$ is clearly the physical light cone ($c = 1$). The equation (27) *per se* will not be used explicitly in the sequel.

Substitutions of the form $N_\lambda = \bar{g}^{\lambda/2}\alpha_\lambda$ ($\lambda > 0$) have also been considered [27]. However, after working out the wave equation analogous to (27) that N_λ obeys, one finds that the local proper propagation speed of N_λ is $\sqrt{\lambda}$. This behavior may or may not spoil the propagation of system variables other than N, but if $\lambda \neq 1$ and the system is hyperbolic, one will always find that the characteristic directions of the system will not all be physical ones. That is, in vacuum gravity, there will be some variables that propagate neither on the light cone (speed $= 1$) nor along the axis parallel to $\bar{\partial}_0$ (speed $= 0$) which is orthogonal to $t =$ const. The variables not propagating in physical directions are gauge variables and one will not have physical criteria for their boundary values on characteristic surfaces. On the other hand, with $\lambda = 1$, one has fulfilled a necessary condition that physical and gauge variables propagate together in the same directions.

In the following sections, whenever we consider hyperbolic systems, we will focus on first-order symmetric (or symmetrizable) hyperbolic ("FOSH") equations possessing *only* physical characteristic directions. We understand "FOSH" in *this restricted physical sense only* in this paper, and likewise for other uses of the term "hyperbolic."

3 Canonical Action and Equations of Motion

Choice of the slicing function α in (25) is arbitrary ($\alpha > 0$). That α is freely chosen while N must satisfy an equation of motion (23) suggests that it should be regarded as the undetermined[2] multiplier in the canonical action principle of Arnowitt, Deser, and Misner ("ADM") [2]. Here we follow [29].

To draw some lessons for the canonical formalism, let us first express the $3+1$ evolution equations in their standard geometrical form (see: with zero shift [8], arbitrary lapse and shift [7], spacetime perspective [30]):

$$\dot{g}_{ij} \equiv -2NK_{ij} \,, \qquad (28)$$

$$\dot{K}_{ij} \equiv N\left(-R_{ij}(g) + R_{ij}(\bar{g}) + NK_{ij} - K_{ik}K^k_j - N^{-1}\bar{\nabla}_i\partial_j N\right) \,, \qquad (29)$$

[2] The multipliers associated with gauge freedom or, as here, with spacetime coordinate freedom, can be freely chosen because they are not determined by physical conditions. Hence, they are not true "Lagrange multipliers."

where $\dot{(\ )} \equiv \bar{\partial}_0(\ )$.

A brief look at (29) shows that forming the combination $\mathcal{R}_{ij} = R_{ij}(g) - g_{ij}R^k{}_k(g)$ leads to an equation of motion for the ADM canonical momentum

$$\pi^{ij} = \bar{g}^{1/2}\left(Kg^{ij} - K^{ij}\right) \tag{30}$$

that contains no constraints. (In this section we choose units in which $16\pi G = c = 1$.) Indeed, using (28) and (29), we obtain the identity

$$\begin{aligned}\dot{\pi}^{ij} \equiv\ & N\bar{g}^{1/2}\left(R(\bar{g})g^{ij} - R^{ij}(\bar{g})\right) - N\bar{g}^{-1/2}\left(2\pi^{ik}\pi^j{}_k - \pi\pi^{ij}\right) \\ & + \bar{g}^{1/2}\left(\bar{\nabla}^i\bar{\nabla}^j N - g^{ij}\bar{\nabla}_k\bar{\nabla}^k N\right) + N\bar{g}^{1/2}\left[\mathcal{R}^{ij}\right]\ .\end{aligned} \tag{31}$$

From the identity $\dot{g}_{ij} = -2NK_{ij}$, we have

$$\dot{g}_{ij} \equiv N\bar{g}^{-1/2}\left(2\pi_{ij} - \pi g_{ij}\right)\ . \tag{32}$$

We now come to a crucial observation. Were the canonical equation for $\dot{\pi}^{ij}$ to be dictated by vanishing of the spatial part of the Einstein tensor, $G^{ij}(g) = 0$, as it is in the conventional ADM analysis [2], then the identity

$$G_{ij}(g) + g_{ij}G^0{}_0(g) \equiv R_{ij}(g) - g_{ij}R^k{}_k(g) \equiv \mathcal{R}_{ij} \tag{33}$$

shows that a Hamiltonian constraint term $\sim \bar{g}^{1/2}G^0{}_0$ remains in the $\dot{\pi}^{ij}$ equation (31). This would mean that the validity of the $\dot{\pi}^{ij}$ equation would be restricted to the subspace on which the Hamiltonian constraint is satisfied (i.e., vanishes).

Though the ADM derivation of the $\dot{\pi}^{ij}$ equation, found by varying g_{ij} in their canonical action (β^i is the shift vector)

$$S\left[g,\pi;N,\beta\right] = \int d^4x\left(\pi^{ij}\dot{g}_{ij} - N\mathcal{H}\right)\ , \tag{34}$$

with $N(t,x)$, $\beta^i(t,x)$ and π^{ij} held fixed, is of course perfectly correct, another point of view is possible. [We are ignoring boundary terms, a subject not of interest here, and we note that the momentum constraint term $-\beta^i\mathcal{H}_i$ ($\mathcal{H}_i = g^{1/2}\mathcal{C}_i$, $\mathcal{C}_i = 2NR^0{}_i$) is contained in $\pi^{ij}\dot{g}_{ij}$ ($\dot{(\ )} \equiv \bar{\partial}_0$) upon integration by parts.] (The slicing density α has also been used prominently in the action by Teitelboim [31], who simply set $\alpha = 1$ ($N = \bar{g}^{1/2}$), and by Ashtekar [32, 33] for other purposes.)

We have explained that α can be regarded as a free undetermined multiplier while N is a *dynamical variable* (a conclusion also reached by Ashtekar for other reasons [32,33]) that determines the proper time $N\delta t$ between slices $t = t'$ and $t = t' + \delta t$. N is determined from $\alpha(t,x)$ and $\bar{g}^{1/2}$ found by solving the initial value constraint equations. (See the treatment of the constraints in the final section of this article and in [30,34,35,9].) Motivated by this viewpoint, we alter the undetermined multiplier N in the ADM action principle to α, where the Hamiltonian density $\tilde{\mathcal{H}}$ is (with $\mathcal{H} \equiv 2\bar{g}^{1/2} G^0{}_0(g)$ being the ADM Hamiltonian density of weight +1)

$$\tilde{\mathcal{H}} \equiv \bar{g}^{1/2}\mathcal{H} = \pi_{ij}\pi^{ij} - \frac{1}{2}\pi^2 - \bar{g}R(\bar{g}) \ , \tag{35}$$

a scalar density of weight +2 and a rational function of the metric. The action becomes

$$S\left[\bar{g}, \pi; \alpha, \beta\right) = \int d^4x \left(\pi^{ij}\dot{g}_{ij} - \alpha\tilde{\mathcal{H}}\right) \ . \tag{36}$$

The modified action *principle* for the canonical equations that we propose in (36) is to vary π^{ij} and g_{ij}, with $\alpha(t,x)$ and $\beta^i(t,x)$ as fixed undetermined multipliers. From

$$\begin{aligned} \delta\tilde{\mathcal{H}} = & \left(2\pi_{ij} - g_{ij}\pi\right)\delta\pi^{ij} + \left(2\pi^{ik}\pi^j{}_k - \pi\pi^{ij} + \bar{g}R^{ij}(\bar{g}) - \bar{g}g^{ij}R(\bar{g})\right)\delta g_{ij} \\ & -\bar{g}\left(\bar{\nabla}^i\bar{\nabla}^j\delta g_{ij} - g^{ij}\bar{\nabla}_k\bar{\nabla}^k\delta g_{ij}\right) \ , \end{aligned} \tag{37}$$

we obtain the canonical equations

$$\dot{g}_{ij} = \alpha\frac{\delta\tilde{\mathcal{H}}}{\delta\pi^{ij}} = \alpha\left(2\pi_{ij} - \pi g_{ij}\right) \equiv -2NK_{ij} \ , \tag{38}$$

$$\begin{aligned} \dot{\pi}^{ij} = -\alpha\frac{\delta\tilde{\mathcal{H}}}{\delta g_{ij}} = & -\alpha\bar{g}\left(R^{ij}(\bar{g}) - R(\bar{g})g^{ij}\right) - \alpha\left(2\pi^{ik}\pi^j{}_k - \pi\pi^{ij}\right) \\ & +\bar{g}\left(\bar{\nabla}^i\bar{\nabla}^j\alpha - \bar{g}^{ij}\bar{\nabla}_k\bar{\nabla}^k\alpha\right) \ . \end{aligned} \tag{39}$$

Equation (39) for $\dot{\pi}^{ij}$ is the identity (31) with $\mathcal{R}^{ij} = 0$, which is equivalent to $R_{ij}(g) = 0$. Thus, (39) is a "strong" equation unlike its ADM counterpart, which requires in addition the imposition of a constraint: $\mathcal{H} = 0$.

In the present formulation, the canonical equations of motion hold everywhere on phase space with any parameter time t, a necessary condition for the issue of "constraint evolution" even to be *discussed* in the Hamiltonian framework. (See below in Sect. 4.)

If we define the "smeared" Hamiltonian as the integral of the Hamiltonian density,

$$\tilde{\mathcal{H}}_\alpha = \int d^3x'\alpha(t,x')\tilde{\mathcal{H}} \ , \tag{40}$$

the equation of motion for a general functional $F[\bar{g}, \pi; t, x)$ anywhere on the phase space is

$$\dot{F}\left[\bar{g}, \pi; t, x\right) = -\left\{\tilde{\mathcal{H}}_\alpha, F\right\} + \tilde{\partial}_0 F \ , \tag{41}$$

where $\dot{(\ )}$ denotes our total time derivative and $\tilde{\partial}_0$ is a "partial" derivative of the form $\partial_t - \mathcal{L}_\beta$ acting only on explicit spacetime dependence. The Poisson bracket is

$$\{F, G\} = \int d^3x \left(\frac{\delta F}{\delta g_{ij}(t,x)}\frac{\delta G}{\delta\pi^{ij}(t,x)} - \frac{\delta G}{\delta g_{ij}(t,x)}\frac{\delta F}{\delta\pi^{ij}(t,x)}\right) \ , \tag{42}$$

and one sees that time evolution is generated by the Hamiltonian vector field

$$\mathcal{X}_{\tilde{\mathcal{H}}_\alpha} = \int d^3x \left\{ \alpha(2\pi_{ij} - \pi g_{ij}) \frac{\delta}{\delta g_{ij}} - [\alpha\bar{g}(R^{ij}(\bar{g}) - R(\bar{g})g^{ij}) \right.$$

$$\left. + \alpha(2\pi^{ik}\pi^j{}_k - \pi\pi^{ij}) - \bar{g}(\bar{\nabla}^i\bar{\nabla}^j\alpha - g^{ij}\bar{\nabla}_k\bar{\nabla}^k\alpha)] \frac{\delta}{\delta\pi^{ij}} \right\} . \quad (43)$$

Because it does not contain any explicit constraint dependence, (43) is a valid time evolution operator on the entire phase space. It is clear that the $(\dot{\bar{g}}, \dot{\pi})$ equations come from (43) applied to the canonical variables. The harmonic time slicing equation (23) results from application of (43) to N, and the wave equation for N comes from a repeated application of (43) to (23).

Evolution equations for the "constraints" are computed to be

$$\bar{\partial}_0\tilde{\mathcal{H}} = -\left\{\tilde{\mathcal{H}}_\alpha, \tilde{\mathcal{H}}\right\} = \alpha\bar{g}g^{ij}\partial_i\mathcal{H}_j + 2\bar{g}g^{ij}\mathcal{H}_i\bar{\nabla}_j\alpha , \quad (44)$$

$$\bar{\partial}_0\mathcal{H}_j = -\left\{\tilde{\mathcal{H}}_\alpha, \mathcal{H}_j\right\} = \alpha\partial_j\tilde{\mathcal{H}} + 2\tilde{\mathcal{H}}\partial_j\alpha , \quad (45)$$

where $\bar{\nabla}_j\alpha = \partial_j\alpha + \alpha\bar{g}^{-1/2}\partial_i\bar{g}^{1/2}$. These are well-posed evolution equations for the constraints, and they are equivalent to the twice-contracted Bianchi identities when $\mathcal{R}_{ij} = 0$ or $R_{ij} = 0$ (see below).

These results shed new light on the Dirac "algebra" of constraints [36]. It is well known that the Dirac algebra is not the spacetime diffeomorphism algebra. This can be seen from the fact that while the action (36) is invariant under transformations generated by $\mathcal{H}_j$ and $\tilde{\mathcal{H}}$, [37] the equations of motion that follow from this action are $R_{ij}(g) = 0$ even when $\mathcal{H}_j$ and $\tilde{\mathcal{H}}$ do not vanish. These equations of motion are preserved by spatial diffeomorphisms and time translations along their flow in phase space, whereas a general spacetime diffeomorphism applied to $R_{ij}(g) = 0$ would mix in the constraints.

A second important view of the Dirac algebra results from the direct and beautiful dynamical meaning of its once-smeared form. Equations (44) and (45) express *consistency* of the constraints as a *well posed* initial-value problem. If the constraint functions vanish in some region on an intial time slice, they continue to do so under evolution by the Hamiltonian vector field into the domain of dependence of that initial region. This mechanism follows from the dual role of $\tilde{\mathcal{H}}$ as a constraint and as part of the generator of time translations of functionals of the canonical variables anywhere on the phase space.

Let us take note that the Hamiltonian constraint *per se* does not express the dynamics of the theory; the equation of dynamics is (41). In its "altered" role, the Hamiltonian constraint function simply vanishes as an initial value condition , from which $\bar{g}^{1/2}$ is determined as in the initial value problem. [30] Then N can be constructed from α. The Hamiltonian constraint, once solved, remains so according to the results embodied in (44) and (45).

4 Contracted Bianchi Identities

The results on canonical dynamics that follow on using α as an undetermined multiplier are also reflected in the manner in which the twice-contracted Bianchi identities,

$$\nabla_\beta G^\beta{}_\alpha \equiv 0 \ , \tag{46}$$

can be written as a first-order symmetrizable hyperbolic system [29]. (In the absence of hyperbolic form, (46) is practically useless in providing physical equations of motion for the constraints when they are *not* satisfied.) Likewise, this system extends to matter (see below). (Frittelli obtained well-posedness for (46) by other methods [38].)

We recall that the equations of motion of the canonical momenta in vacuum are

$$\mathcal{R}_{ij} \equiv R_{ij}(g) - g_{ij}R^k{}_k(g) = 0 \ , \tag{47}$$

while the weight zero Hamiltonian constraint is

$$C = 2G^0{}_0(g) = K_{ij}K^{ij} - K^2 - R(\bar{g}) = 0 \ , \tag{48}$$

and the weight zero one-form momentum constraint is

$$C_i = 2NR^0{}_i(g) = 2\bar{\nabla}^j(K_{ij} - Kg_{ij}) = 0 \ . \tag{49}$$

Recall the identity (33):

$$G_{ij}(g) + g_{ij}G^0{}_0(g) \equiv R_{ij}(g) - g_{ij}R^k{}_k(g) \equiv \mathcal{R}_{ij} \ . \tag{50}$$

Combining (47), (48), (49), and (33) with (46) gives the twice-contracted Bianchi identities as a FOSH system

$$\dot{C} - N\bar{\nabla}^j C_j \equiv 2\left(C_j\bar{\nabla}^j N + NKC - NK^{ij}[\mathcal{R}_{ij}]\right) \ , \tag{51}$$

$$\dot{C}_j - N\bar{\nabla}_j C \equiv 2\left(C\bar{\nabla}_j N + \frac{1}{2}NKC_j - \bar{\nabla}^i(N[\mathcal{R}_{ij}])\right) \ . \tag{52}$$

Substituting $\mathcal{H}_i = \bar{g}^{1/2}C_i$, $\tilde{\mathcal{H}} = \bar{g}C$, and setting the equations of motion $\mathcal{R}_{ij} = 0$ in (51) and (52) yields the evolution equations of the unsmeared constraints as in (44) and (45).

Similar considerations show how to put the "matter conservation" equations $\nabla_\beta T^{\alpha\beta} = 0$ into well-posed form. This was also carried out by one of us (YCB) and Noutchegueme [39] but the results obtained here are more immediately physical. Unlike [39], we use the energy density $\varepsilon = -T^0{}_0$ rather than $\rho^{00}(\rho^\alpha{}_\beta \equiv T^\alpha{}_\beta - \frac{1}{2}\delta^\alpha{}_\beta T^\mu{}_\mu)$ to obtain this result. It is clear that such a result is possible because

$$H_{\alpha\beta} \equiv \kappa^{-1}G_{\alpha\beta}(g) - T_{\alpha\beta} \tag{53}$$

vanishes as Einstein's equation and in any case satisfies $\nabla_\beta H^\beta{}_\alpha = 0$. We can treat $H_{\alpha\beta}$ as we did $G_{\alpha\beta}$ above. ($\kappa = 8\pi G$; $c = 1$.) The result is nevertheless of interest as it presents the continuity and relativistic Euler equations of matter in a well-posed form.

Straightforwardly expanding $\nabla_\beta T^\beta{}_0 = 0$ and $\nabla_\beta T^\beta{}_i = 0$ gives the continuity and Euler equations (cf. [30], p. 89), with $\varepsilon \equiv -T^0{}_0$ and the matter current one-form $j_i \equiv NT^0{}_i$. The continuity equation is

$$\bar{\partial}_0 \varepsilon + N\bar{\nabla}^i j_i = N(K_{ij}T^{ij} + K\varepsilon - 2j_i a^i) \; , \tag{54}$$

where $a_i \equiv \bar{\nabla}_i \log N$ is the acceleration of observers at rest in a given time-slice. Likewise, we find for Euler's equation

$$\bar{\partial}_0 j_i + N\bar{\nabla}_j T^j{}_i = N(K^j{}_i - T^j{}_i a_j - \varepsilon a_i) \; . \tag{55}$$

The divergence term on the left side of (55) spoils the well-posed FOSH form we seek. However, if we use the identity $G_{ij}(g) + g_{ij}G^0{}_0(g) \equiv R_{ij}(g) - g_{ij}R^k{}_k(g)$ and the Einstein equations $\kappa^{-1}G_{\alpha\beta} - T_{\alpha\beta} = 0$, or $\kappa^{-1}R_{\alpha\beta} = \rho_{\alpha\beta}$, we obtain ($\varepsilon = -T^0{}_0$, $j_i = NT^0{}_i$)

$$T^j{}_i - \delta^j{}_i\varepsilon = (\rho^j{}_i - \delta^j{}_i\rho^k{}_k) \equiv \mathcal{S}^j{}_i \; . \tag{56}$$

Then (54) and (55) obtain well-posed form (if $\mathcal{S}^j{}_i$ is assumed known),

$$\bar{\partial}_0 \varepsilon + N\bar{\nabla}^i j_i = -2j_i\bar{\nabla}^i N + 2NK\varepsilon + NK^{ij}\mathcal{S}_{ij} \; , \tag{57}$$

$$\bar{\partial}_0 j_i + N\bar{\nabla}_i \varepsilon = -2\varepsilon\bar{\nabla}_i N + NKj_i - \bar{\nabla}^j(N\mathcal{S}_{ij}) \; . \tag{58}$$

By combining (57) plus (49), and (58) plus (50), we obtain expressions of gravity constraint evolution in the presence of matter,

$$\bar{\partial}_0 C^T - N\bar{\nabla}^i C^T_i = 2\left(C^T_j\bar{\nabla}^j N + NKC^T - NK^{ij}(\kappa^{-1}\mathcal{R}_{ij} - \mathcal{S}_{ij})\right) \; , \tag{59}$$

$$\bar{\partial}_0 C^T_j - N\bar{\nabla}_j C^T = 2\left[C^T\bar{\nabla}_j N + \frac{1}{2}NKC^T_j - \bar{\nabla}^i\left(N(\kappa^{-1}\mathcal{R}_{ij} - \mathcal{S}_{ij})\right)\right] \tag{60}$$

where $C^T \equiv C + 2\varepsilon$ and $C^T_j = C_j - 2j_j$. This is just the form we would anticipate on the basis of Hamiltonian dynamics and the form (49) and (50) of the vacuum constraints. Thus, for gravity plus a matter field, we obtain results analogous to (44) and (45) for the total system. If there are no violations of constraints, then $C^T = 0$, $C^T_j = 0$, while the dynamical gravity equation is $\kappa^{-1}\mathcal{R}_{ij} - \mathcal{S}_{ij} = 0$.

5 Wave Equation for K_{ij}

Einstein's equations, viewed mathematically as a system of second-order partial differential equations for the metric, do not form a hyperbolic system without modification and are not manifestly well-posed, though, of course, physical information does propagate at the speed of light. A well-posed hyperbolic system admits unique solutions depending continuously on the initial data and seems to be required for robust, stable numerical integration and for full treatment by the methods of modern analysis, for example, exploitation of energy estimates. The well-known traditional approach achieves hyperbolicity through special coordinate choices.[3] The formulation described here permits coordinate gauge freedom. Because these exact nonlinear theories incorporate the constraints, they are natural starting points for developing gauge-invariant perturbation theory.

Consider a globally hyperbolic manifold $V = \Sigma \times \mathbb{R}$ with the metric as given in the introduction. To achieve hyperbolicity for the $3+1$ equations, we proceed as follows.

By taking a time derivative of $R_{ij}(g)$ and subtracting appropriate spatial covariant derivatives of the momentum constraints, one of us (YCB) and T. Ruggeri [18] (see also [19], where the shift is not set to zero) obtained an equation with a wave operator acting on the extrinsic curvature. In vacuum, one finds

$$\bar{\partial}_0 R_{ij}(g) - \bar{\nabla}_i R_{0j} - \bar{\nabla}_j R_{0i} = N \bar{\Box}_g K_{ij} + J_{ij} + S_{ij} = 0 \,, \tag{61}$$

where $\bar{\Box}_g = -\left(N^{-1}\bar{\partial}_0\right)^2 + \bar{\nabla}_k \bar{\nabla}^k$, J_{ij} consists of terms at most first order in derivatives of K_{ij}, second order in derivatives of g_{ij}, and second order in derivatives of N, and

$$S_{ij} = -N^{-1} \bar{\nabla}_i \bar{\nabla}_j (\bar{\partial}_0 N + N^2 K) \,. \tag{62}$$

The term S_{ij} is second order in derivatives of K_{ij} and would spoil hyperbolicity of the wave operator $\bar{\Box}$ acting on K_{ij}. Hyperbolicity is achieved by setting $N = \bar{g}^{1/2}\alpha(t,x)$, or

$$\bar{\partial}_0 N + N^2 K = \bar{g}^{1/2} \bar{\partial}_0 \alpha(t,x) \,, \tag{63}$$

as discussed in Sect. 2. The resulting equation combined with (28) forms a quasi-diagonal hyperbolic system for the metric g_{ij} with principal operator $\bar{\partial}_0 \bar{\Box}$. This system can also be put in first order symmetric hyperbolic form [19,11], by the introduction of sufficient auxiliary variables and by use of the equation for R_{00} (thus incorporating the Hamiltonian constraint). The Cauchy data for the system (in vacuum) [18,19] are (1) $(\bar{g}, K)$ such that the constraints $R_{0i} = 0$, $G^0{}_0 = 0$ hold on the initial slice; (2) $\bar{\partial}_0 K_{ij}$ such that $R_{ij} = 0$ on the intial slice; and (3) $N > 0$ arbitrary on the initial slice. Note that the shift $\beta^k(x,t)$ is arbitrary. Using the Bianchi identities, one

[3] The classic second-order fully harmonic form was given in [6,7] and discussed, for example, in [9]. It will not be discussed in this article. A FOSH form based on these equations was given first by Fischer and Marsden in [23].

can prove [18,19] that this system is fully equivalent to the Einstein equations. The point is that quasi-diagonal Leray [40] hyperbolic systems have well posed Cauchy problems and therefore unique solutions for given initial data. Because every solution of the Einstein equations also satisfies the $\bar{\Box} K_{ij}$ equation in particular and provides initial data for it, uniqueness implies, conversely, that if the initial data for the $\bar{\Box} K_{ij}$ equation are Einsteinian, all solutions of Einstein's equations, and only these, are captured. The restriction on the initial value of $\bar{\partial}_0 K_{ij}$ prevents the higher derivative from introducing spurious unphysical solutions.

All variables propagate either with characteristic speed zero or the speed of light. The only variables which propagate at the speed of light have the dimensions of curvature, and one sees that this is a theory of propagating curvature. However, a FOSH system that propagates curvature is more transparent in the "Einstein-Bianchi" form (next section).

In the above formulation, the shift and α are arbitrary. This and our other systems (except the Einstein-Christoffel system in Sect. 7) are manifestly spatially covariant and all time slicings (using α) are allowed. Spacetime covariance is therefore present, but not completely manifest.

By taking another time derivative and adding an appropriate derivative of R_{00}, one finds (in vacuum) [29]

$$\bar{\partial}_0 \bar{\partial}_0 R_{ij} - \bar{\partial}_0 \bar{\nabla}_i R_{0j} + \bar{\partial}_0 \bar{\nabla}_j R_{0i} + \bar{\nabla}_i \bar{\nabla}_j R_{00} = \bar{\partial}_0 (N \bar{\Box} K_{ij}) + \mathcal{J}_{ij} = 0 \;, \quad (64)$$

where $\mathcal{J}_{ij}$ consists of terms at most third order in derivatives of g_{ij} and second order in derivatives of K_{ij}. Together with $\bar{\partial}_0 g_{ij}$, these form a system for $(\bar{g}, K)$ which is hyperbolic non-strict in the sense of Leray-Ohya. [41] Here, the lapse itself, as well as the shift, is arbitrary ($N > 0$). The Cauchy data of the previous form (in vacuum) must be supplemented by $\bar{\partial}_0 \bar{\partial}_0 K_{ij}$ such that $\bar{\partial}_0 R_{ij} = 0$ on the initial slice. This guarantees that the system is fully equivalent to Einstein's theory (except that its solutions are not in Sobolev spaces [29]). This system does not have a first order symmetric hyperbolic formulation, but has been used very effectively in perturbation theory [42] and in other applications [43,44].

6 Einstein-Bianchi Hyperbolic System

To obtain a first order symmetric hyperbolic system, one can use the Riemann tensor of the spacetime metric. It satisfies the Bianchi identities for the spacetime geometry

$$\nabla_\alpha R_{\beta\gamma\lambda\mu} + \nabla_\beta R_{\gamma\alpha\lambda\mu} + \nabla_\gamma R_{\alpha\beta\lambda\mu} \equiv 0 \;. \quad (65)$$

These identities imply by contraction and use of the symmetries of the Riemann tensor

$$\nabla_\alpha R^\alpha{}_{\mu\beta\gamma} + \nabla_\gamma R_{\beta\mu} + \nabla_\beta R_{\gamma\mu} \equiv 0 \;. \quad (66)$$

If the Ricci tensor $R_{\alpha\beta}$ satisfies the Einstein equations ($\kappa = c = 1$)

$$R_{\alpha\beta} = \rho_{\alpha\beta} \ , \tag{67}$$

then the previous identities imply the equations

$$\nabla_\alpha R^\alpha{}_{\mu\beta\gamma} = \nabla_\beta \rho_{\gamma\mu} - \nabla_\gamma \rho_{\beta\mu} \ . \tag{68}$$

The first equations with $(\alpha\beta\gamma) = (ijk)$ and the last one with $\mu = 0$ do not contain derivatives of the Riemann tensor transverse to M_t. They are considered as "constraints" and will be identically satisfied (initially) in our method. They remain satisfied in an exact integration. All detail and rigor concerning this elegant system is given in [22,51], to which the reader is referred. It has 66 equations, just as do the Einstein-Ricci first order curvature equations.

The system we are now developing [21] is similar to an analogous system obtained by H. Friedrich [25] that is based on the Weyl tensor. The Weyl tensor system is causal but with additional unphysical characteristics.

We wish first to show that the remaining equations are, for $n = 3$ in the vacuum case, when g is given, a symmetric first order hyperbolic system for the double two-form $R_{\alpha\beta\lambda\mu}$. For this purpose, following Bel [52,53] we introduce two pairs of "electric" and "magnetic" space tensors associated with a spacetime double two-form A,

$$N^2 E_{ij}(g) \equiv A_{0i0j} \tag{69}$$

$$D_{ij}(g) \equiv \frac{1}{4} \epsilon_{ihk} \epsilon_{jlm} A^{hklm} \tag{70}$$

$$N H_{ij}(g) \equiv \frac{1}{2} \epsilon_{ihk} A^{hk}{}_{0j} \tag{71}$$

$$N B_{ji}(g) \equiv \frac{1}{2} A_{0j}{}^{hk} \epsilon_{ihk} \tag{72}$$

where ϵ_{ijk} is the volume form of $\bar{g}$. It results from the symmetry of the Riemann tensor R with respect to its first and second pairs of indices (R is a "symmetric double two-form") that if $A \equiv R$, then E and D are symmetric while $H_{ij} = B_{ji}$. A useful identity for a symmetric double two-form like R, with a tilde representing the spacetime double dual, is ("Lanczos identity")

$$\tilde{R}_{\alpha\beta\lambda\mu} + R_{\alpha\beta\lambda\mu} = C_{\alpha\lambda}\, g_{\beta\mu} - C_{\alpha\mu}\, g_{\beta\lambda} + C_{\beta\mu}\, g_{\alpha\lambda} - C_{\beta\lambda}\, g_{\alpha\mu} \ , \tag{73}$$

where $C_{\alpha\beta} = R_{\alpha\beta} - (1/4) g_{\alpha\beta} R$. It follows that when $R_{\alpha\beta} = \lambda\, g_{\alpha\beta}$, then $E = -D$ and $H = B$. In order to avoid introducing unphysical characteristics, and to be able to extend the treatment to the non-vacuum case, *we do not use these properties in the evolution equations*, but write them as a first order system for an arbitrary double two-form A, as follows:

$$\nabla_0 A_{hk0j} + \nabla_k A_{0h0j} - \nabla_h A_{0k0j} = 0 \ , \tag{74}$$

$$\nabla_0 A^0{}_{i0j} + \nabla_h A^h_{i0j} = \nabla_0\, \rho_{ji} - \nabla_j\, \rho_{0i} \ , \tag{75}$$

and analogous equations with the pair $(0\,j)$ replaced by $(l\,m)$. One obtains a first order system for the unknowns E, H, D, and B by using the relations inverse to the definitions above. The principal parts of these equations, all with one definite index fixed on E, H, D, and B, are identical to the corresponding Maxwell equations. The characteristic matrix of this "Maxwell" part of the system has determinant $-N^6(\xi_0\xi^0)\,(\xi_\alpha\xi^\alpha)^2$. The system obtained has a principal matrix consisting of 6 identical 6 by 6 blocks around the diagonal, which are symmetrizable and hyperbolic. Hence, the system is symmetric hyperbolic, when g is a given metric such that $\bar{g}$ is properly riemannian and $N > 0$.

To relate the Riemann tensor to the metric $\bar{g}$ we use the definition

$$\bar{\partial}_0 g_{ij} = -2NK_{ij} \tag{76}$$

and we use the 3 + 1 identities given in the Introduction. *Note that in this section all Γ's are spatial.*

We next choose $N = \bar{g}^{1/2}\alpha(t,x)$. We generalize somewhat the ideas used by Friedrich (see [25]) for the Weyl tensor to write a symmetric hyperbolic system for K and Γ, namely we obtain equations relating Γ and K, for a given double two-form A, and by considering the definition of K and the 3 + 1 decomposition of the Riemann tensor, replacing in these identities the Riemann tensor by A. To deduce from this system a symmetric hyperbolic first order system, with the algebraic form of the harmonic gauge $N = \bar{g}^{1/2}\alpha$, one uses the fact that in this gauge one has

$$\Gamma^h{}_{ih} = \partial_i \log N - \partial_i \log \alpha\;. \tag{77}$$

We obtain

$$\begin{aligned}\bar{\partial}_0 \Gamma^h{}_{ij} + N\bar{\nabla}^h K_{ij} = {} & NK_{ij}g^{hk}\left(\Gamma^m{}_{mk} + \partial_k \log\alpha\right) \\ & -2NK^h{}_{(i}\left(\Gamma^m{}_{j)m} + \partial_{j)}\log\alpha\right) \\ & -N\left(\epsilon^k{}_{(j}{}^h B_{i)k} + H_{k(i}\epsilon^k{}_{j)}{}^h\right)\;,\end{aligned} \tag{78}$$

and

$$\begin{aligned}\bar{\partial}_0 K_{ij} + N\partial_h \Gamma^k{}_{ij} = {} & N\left[\Gamma^m{}_{ih}\Gamma^h{}_{jm} - \left(\Gamma^h{}_{ih} + \partial_i \log\alpha\right)\left(\Gamma^k{}_{jk} + \partial_j\log\alpha\right)\right] \\ & -N\left(\partial_i\partial_j \log\alpha - \Gamma^k{}_{ij}\partial_k \log\alpha\right) \\ & -N\left[D_{(ij)} + E_{(ij)} - D^k{}_k g_{ij} - KK_{ij}\right]\;.\end{aligned} \tag{79}$$

The system obtained for K and Γ has a characteristic matrix composed of 6 blocks around the diagonal, each block a 4 by 4 matrix that is symmetrizable hyperbolic, with characteristic polynomial $-N^4(\xi_0\xi^0)\,(\xi_\alpha\xi^\alpha)$.

The whole system for A, K, Γ, $\bar{g}$ is symmetrizable hyperbolic, with characteristics the light cone and the normal to M_t. It is somewhat involved to prove that a solution of the constructed system satisfies the Einstein equations if the initial data satisfy the constraints, but we can argue as follows.

We consider the vacuum case with initial data satisfying the Einstein constraints. These initial data determine the initial values of Γ, and also, if β and N are known at $t = 0$, the initial values of A_{ijhm}, A_{jhi0}, A_{i0jh} by using the decomposition formulas. (We set A equal to the Riemann tensor on the initial surface.) We use the Lanczos formula to determine A_{i0j0} initially. We know that our symmetrizable hyperbolic system has one and only one solution. Because a solution of Einstein's equations with $N = \bar{g}^{1/2}\alpha$, proved to exist in the section on $\bar{\Box}K_{ij}$, satisfies together with its Riemann tensor the present system and takes the same initial values, that solution coincides with the solution of the present system in their common domain of existence.

7 Einstein-Christoffel System

The first-order form of the wave equation for K_{ij}, the Einstein-Ricci system [19,11], to which we have alluded, has 66 equations, the correct number for a curvature system as does the Einstein-Bianchi system of the previous section. It is symmetric hyperbolic and therefore well-posed. But it is natural to ask whether there is a simpler first-order system of fewer variables that is perhaps closer in form to (28) and (29). Frittelli and Reula [26,27] proposed such a system, but their system has unphysical characteristics and is not written fully in terms of geometric variables. Here we deduce a different system having only physical characteristics and expressed in geometric variables. We understand (private communication to JWY) that James Bardeen has likewise obtained a similar system improving that found in [16,17]. While our derivation [10] can proceed systematically by direct construction of an energy norm and the characteristic speeds, a more heuristic derivation is indicated here from the structure of the wave equation for K_{ij}. Not that in this section all Γ's are spatial.

In the dynamical spacetime Ricci tensor $R_{ij}(g)$, (29), one has a dynamical equation for the extrinsic curvature K_{ij} in terms of spatial derivatives of the spatial Christoffel symbols. When the lapse N is replaced by the slicing density through $\alpha\bar{g}^{1/2}$, the differentiated Christoffel terms become

$$\partial_k \Gamma^k{}_{ij} - \partial_j \Gamma^k{}_{ik} - \partial_i \Gamma^k{}_{jk} \ . \tag{80}$$

This may be read as the divergence of a linear combination of Christoffel symbols, which puts the $R_{ij}(\bar{g})$ equation in a form reminiscent of the structure of one of the first-order equations for a free wave, namely

$$\partial_0 u + \partial^k v_k = 0 \ . \tag{81}$$

We would have a symmetric hyperbolic system if there were an analog of the other equation for a wave,

$$\partial_0 v_k + \partial_k u = 0 \ . \tag{82}$$

Some manipulation quickly leads to the conclusion that K_{ij} and the Christoffel combination above are not paired in a symmetric hyperbolic system like u and v_k.

Recall however, that the free wave equation $(\partial_0)^2 u - \partial^k \partial_k u = 0$ is obtained by taking a time derivative of (81) and subtracting the divergence of (82). In obtaining the wave equation for K_{ij} (61) we have taken a time derivative of R_{ij} and subtracted a (suitably symmetrized) divergence of the momentum constraint R_{0i}. This motivates the speculation that in gravity the "other" equation should be related to the momentum constraint and its sole spatial derivative should be $\partial_k K_{ij}$.

Following this idea about the second equation leads one to consider an equation of the form

$$g_{ki} R_{j0} + g_{kj} R_{i0} = -\bar{\partial}_0 f_{kij} - \partial_k (N K_{ij}) + l.o._{kij} \tag{83}$$

where $l.o._{kij}$ are lower order terms involving no derivatives of f_{kij} or K_{ij}. One must choose f_{kij} from a linear combination of spatial derivatives of the metric. Introduce

$$\mathcal{G}_{kij} = \partial_k g_{ij} \tag{84}$$

and use the identity

$$\bar{\partial}_0 (\partial_k g_{ij}) = -\partial_k (2 N K_{ij}) \tag{85}$$

to find that

$$f_{kij} = \frac{1}{2} \mathcal{G}_{kij} - g_{k(i} g^{rs} \left(\mathcal{G}_{|rs|j)} - \mathcal{G}_{j)rs} \right) \tag{86}$$

produces the correct coordinate derivatives occuring in the momentum constraints. The lower order terms are those terms necessary to complete (83) into an identity and take the form

$$\begin{aligned} l.o._{kij} = {} & 2 N K_{k(i} g^{rs} \left(\mathcal{G}_{|rs|j)} - \mathcal{G}_{j)rs} \right) \\ & + 2 g_{k(i} \Big[K_{j)m} \partial^m N - K \partial_{j)} N \\ & + N K_{j)m} g^{rs} \Gamma^m{}_{rs}(\mathcal{G}) + \frac{1}{2} N \left(\mathcal{G}_{j)rs} - 2 \mathcal{G}_{|rs|j)} \right) K^{rs} \Big] , \end{aligned} \tag{87}$$

where the spatial Christoffel symbols are constructed from $\mathcal{G}_{ijk}$,

$$\Gamma_{kij}(\mathcal{G}) \equiv (1/2) \left(\mathcal{G}_{jki} + \mathcal{G}_{ikj} - \mathcal{G}_{kij} \right) . \tag{88}$$

(It is clear that only one of the three-index symbols $\mathcal{G}_{kij}$, Γ_{kij}, and f_{kij} is needed, say f_{kij}. The necessary algebra will not be reproduced here.)

One then easily verifies that by expressing (88) in terms of derivatives of the metric (assuming a metric compatible connection), we can manipulate it to take the form of a divergence of f_{kij} plus lower order terms. The dynamical Ricci equation becomes

$$R_{ij} = -N^{-1} \bar{\partial}_0 K_{ij} - \partial^k f_{kij} + l.o._{ij} , \tag{89}$$

where

$$\begin{aligned} l.o._{ij} = KK_{ij} - 2K_{ik}K^k{}_j - \alpha^{-1}\left(\partial_i\partial_j - \Gamma^k{}_{ij}(\mathcal{G})\partial_k\right)\alpha \\ -\left(\Gamma^k{}_{ki}(\mathcal{G}) + \alpha^{-1}\partial_i\alpha\right)\left(\Gamma^m{}_{mj}(\mathcal{G}) + \alpha^{-1}\partial_j\alpha\right) \\ +2\Gamma^k{}_{mk}(\mathcal{G})\Gamma^m{}_{ij}(\mathcal{G}) - \Gamma^k{}_{mj}(\mathcal{G})\Gamma^m_{ik}(\mathcal{G}) \\ +g^{kr}g^{sm}\left[\mathcal{G}_{krs}f_{mij} + \mathcal{G}_{km(i}\mathcal{G}_{j)rs} - \mathcal{G}_{krs}\mathcal{G}_{(ij)m}\right] \ . \end{aligned} \tag{90}$$

Together with (28), (83) and (89) constitute a symmetric hyperbolic system for the evolution of g_{ij}, K_{ij} and f_{kij}.

Note that once f_{kij} or, equivalently, $\mathcal{G}_{kij}$ are introduced as variables, the relation (88) becomes an initial condition and does not *a priori* hold for all time. Equation (83) can be related to metric compatibility by putting it in the form

$$\begin{aligned} 4g_{k(i}R_{j)0} = \bar{\partial}_0\mathcal{G}_{kij} + \partial_k\left(2NK_{ij}\right) \\ -4g_{k(i}N\bar{\nabla}^m\left(K_{j)m} - g_{j)m}K\right) \ . \end{aligned} \tag{91}$$

Here, one sees that if the momentum constraint is satisfied for all time, then

$$\mathcal{G}_{kij} = 2\Gamma_{(ij)k} = \partial_k g_{ij} \tag{92}$$

and the connection is metric compatible. If the momentum constraint is violated, metric compatibility is sacrificed. This shows the price paid to achieve a symmetric hyperbolic system that is close to the canonical equations: the momentum constraints become dynamical, and metric compatibility is lost if the latter are violated.

8 Conformal "Thin Sandwich" Data for the Initial Value Problem

The standard approach to the initial value problem is the "conformal method," the fundamental rudiments of which were introduced by Lichnerowicz [8]. The essentially complete form was developed by two of us (YCB and JWY), see [34,9]. Basic theorems were obtained by us and by O'Murchadha [45,35,46], and Isenberg and Moncrief [47]. The older approach concentrates (in the vacuum case to which we restrict ourselves) on the construction of the spatial metric $g_{ij} = \psi^4\gamma_{ij}$ and the traceless part $A^{ij} = \psi^{-10}\lambda^{ij}$ of the extrinsic curvature, where $K_{ij} = A_{ij} + \frac{1}{3}g_{ij}K$. Here, γ_{ij} is a proper riemannian metric given freely, λ^{ij} is constructed by a tensor decomposition method [34, 48], and K is given freely (not conformally transformed). Note that one may as well assume $\det(\gamma_{ij}) = 1$ because only the conformal equivalence class of the metric matters: the entire method is "conformally covariant."

N.B.: In Sect. 8, only spatial metrics will be used. Therefore, *all overbars are dropped in this final section.*

Here, we discuss a new interpretation of the four Einstein vacuum initial-value constraints. (The presence of matter would add nothing new to the analysis.) Partly in the spirit of a "thin sandwich" viewpoint, this approach is based on prescribing the *conformal* metric [1] on each of two nearby spacelike hypersurfaces ("time slices" $t = t'$ and $t = t' + \delta t$) that make a "thin sandwich" (TS). *Essential use is made of the understanding of the slicing function in general relativity.* The new formulation could prove useful both conceptually and in practice, as a way to construct initial data in which one has a hold on the input data different from that in the currently accepted approach. The new approach allows us to *derive* from its dynamical and metrical foundations the important scaling law $A^{ij} = \psi^{-10}\lambda^{ij}$ for the traceless part of the extrinsic curvature. This rule is simply postulated in the one-hypersurface approach.

The constraint equations on Σ are, in vacuum,

$$\nabla_j(K^{ij} - Kg^{ij}) = 0 \ , \tag{93}$$

$$R(g) - K_{ij}K^{ij} + K^2 = 0 \tag{94}$$

where $R(g)$ is the spatial scalar curvature of g_{ij}, ∇_j is the Levi-Civita connection of g_{ij}; and K is the trace of K_{ij}, also called the "mean curvature" of the slice.

The time derivative of the spatial metric g_{ij} is related to K_{ij}, N, and the shift vector β^i by

$$\partial_t g_{ij} \equiv -2NK_{ij} + (\nabla_i\beta_j + \nabla_j\beta_i) \ , \tag{95}$$

where $\beta_j = g_{ji}\beta^i$. The fixed spatial coordinates x of a point on the "second" hypersurface, as evaluated on the "first" hypersurface, are displaced by $\beta^i(x)\delta t$ with respect to those on the first hypersurface, with an orthogonal link from the first to the second surface as a fiducial reference: $\beta_i = \frac{\partial}{\partial t} * \frac{\partial}{\partial x^i}$, where $*$ is the physical spacetime inner product of the indicated natural basis four-vectors. The essentially arbitrary direction of $\frac{\partial}{\partial t}$ is why $N(x)$ and $\beta^i(x)$ appear in the TS formulation. In contrast, the tensor K_{ij} is always determined by the behavior of the unit normal on one slice and therefore does not possess the kinematical freedom, i.e., the gauge variance, of $\frac{\partial}{\partial t}$. Therefore, N and β^i do not appear in the one-hypersurface IVP for (Σ, g, K).

Turning now to the conformal metrics in the IVP, we recall that two metrics g_{ij} and γ_{ij} are conformally equivalent if and only if there is a scalar $\psi > 0$ such that $g_{ij} = \psi^4\gamma_{ij}$. The conformally invariant representative of the entire conformal equivalence class, in three dimensions, is the weight $(-2/3)$ unit-determinant "conformal metric" $\hat{g}_{ij} = g^{-1/3}g_{ij} = \gamma^{-1/3}\gamma_{ij}$ with $g =$

$\det(g_{ij})$ and $\gamma = \det(\gamma_{ij})$. Note particularly that for any small perturbation, $g^{ij}\delta\hat{g}ij = 0$. We will use the important relation

$$g^{ij}\partial_t\hat{g}_{ij} = \gamma^{ij}\partial_t\hat{g}_{ij} = \hat{g}^{ij}\partial_t\hat{g}_{ij} = 0\ . \tag{96}$$

In the following, rather than use the mathematical apparatus associated with conformally weighted objects such as $\hat{g}_{ij}$, we find it simpler to use ordinary scalars and tensors to the same effect. Thus, let the role of $\hat{g}_{ij}$ on the first surface be played by a given metric γ_{ij} such that the physical metric that satisfies the constraints is $g_{ij} = \psi^4\gamma_{ij}$ for some scalar $\psi > 0$. (This corresponds to "dressing" the initial unimodular conformal metric $\hat{g}_{ij}$ with the correct determinant factor $g^{1/3} = \psi^4\gamma^{1/3}$. This process does not alter the conformal equivalence class of the metric.) The role of the conformal metric on the second surface is played by the metric $\gamma'_{ij} = \gamma_{ij} + u_{ij}\delta t$, where, in keeping with (96), the velocity tensor $u_{ij} = \partial_t\gamma_{ij}$ is chosen such that

$$\gamma^{ij}u_{ij} = \gamma^{ij}\partial_t\gamma_{ij} = 0\ . \tag{97}$$

Then, to first order in δt, γ'_{ij} and γ_{ij} have equal determinants, as desired; but γ_{ij} and γ'_{ij} are not in the same conformal equivalence class in general.

We now examine the relation between the covariant derivative operators D_i of γ_{ij} and ∇_i of g_{ij}. The relation is determined by

$$\Gamma^i{}_{jk}(g) = \Gamma^i{}_{jk}(\gamma) + 2\psi^{-1}\left(2\delta^i{}_{(j}\partial_{k)}\psi - \gamma^{il}\gamma_{jk}\partial_l\psi\right)\ , \tag{98}$$

from which follows the scalar curvature relation first used in an initial-value problem by Lichnerowicz [5],

$$R(g) = \psi^{-4}R(\gamma) - 8\psi^{-5}\triangle_\gamma\psi\ , \tag{99}$$

where $\triangle_\gamma \equiv \gamma^{kl}D_kD_l\psi$ is the scalar Laplacian associated with γ_{ij}.

Next, we solve (95) for its traceless part

$$\partial_t g_{ij} - \frac{1}{3}g_{ij}g^{kl}\partial_t g_{kl} \equiv V_{ij} = -2NA_{ij} + (L_g\beta)_{ij} \tag{100}$$

with $A_{ij} \equiv K_{ij} - (1/3)Kg_{ij}$ and

$$(L_g\beta)_{ij} \equiv \nabla_i\beta_j + \nabla_j\beta_i - (2/3)g_{ij}\nabla^k\beta_k\ . \tag{101}$$

Expression (101) vanishes, for non-vanishing β^i, if and only if g_{ij} admits a conformal Killing vector $\beta^i = k^i$. Clearly, k^i would also be a conformal Killing vector of γ_{ij}, or of any metric conformally equivalent to g_{ij}, with no scaling of k^i. This teaches us that β^i does not scale. That β^i does not scale also follows because, as generator of a spatial diffeomorphism, it is not a dynamical variable. We take the latter "rule" as a matter of principle.

It is clear in (100) that the left hand side u_{ij} satisfies $u_{ij} = \psi^4 V_{ij}$ because the terms in $\dot{\psi}$ cancel out. Furthermore, a straightforward calculation shows that

$$(L_g\beta)_{ij} = \psi^4 \left[L_\gamma(\psi^{-4}\beta)\right]_{ij} \; ; \quad (L_g\beta)^{ij} = \psi^4(L_\gamma\beta)^{ij} \; , \tag{102}$$

where $\psi^{-4}\beta_j = \gamma_{ij}\beta^i$. Next, we note that the lapse function N has essential non-trivial conformal behavior. This is a new element in the IVP analysis. The slicing function $\alpha(t,x) > 0$ can replace the lapse function N,

$$N = g^{1/2}\alpha \; . \tag{103}$$

(The treatment here extends the one in [49] in a simple but interesting way.) We have concluded that α is not a dynamical variable and therefore does not scale. Furthermore, without loss of generality, we can set $\det(\gamma_{ij}) = 1$ and thus $g^{1/2} = \psi^6$. Then

$$N = \psi^6\alpha \; . \tag{104}$$

Finally, we fix K and require that it does not scale, as in the standard treatment of the IVP [50]. This step is absolutely essential for geometric consistency, as we shall see.

Next we solve (98) for A^{ij}, using the scaling rules established above, and find

$$A^{ij} = \psi^{-10} \left\{ \frac{1}{2\alpha} \left[(L_\gamma\beta)^{ij} - u^{ij}\right] \right\} \; . \tag{105}$$

The momentum constraint becomes

$$D_j \left[\frac{1}{2\alpha}(L_\gamma\beta)^{ij}\right] = D_j \left[\frac{1}{2\alpha}u^{ij}\right] + \frac{2}{3}\psi^6\gamma^{ij}\partial_j K \; , \tag{106}$$

while the Hamiltonian constraint becomes [34]

$$8\triangle_\gamma\psi - R(\gamma)\psi + (\gamma_{ik}\gamma_{jl})A^{ij}A^{kl}\psi^{-7} - (2/3)K^2\psi^5 = 0 \; . \tag{107}$$

The unknowns (ψ, β) obey equations of the same form as do the conformal scalar potential ϕ and the vector potential W^i in the standard analysis [34, 9], but *no tensor splittings* are required. Further, (106) and (107) are coupled in only one direction when $K = \text{const}$.

Now we note two interesting consequences of this approach. First we see that from $N = \psi^6\alpha$, we have identically

$$N = g^{1/2}\alpha \tag{108}$$

as a consequence of the method. Therefore, time slices t and $t + \delta t$ have a relation that is manifestly "harmonic:"

$$\bar{\partial}_0 N + N^2 K = N\bar{\partial}_0 \log\alpha \; , \tag{109}$$

a result that is fully consistent with our previous discussions and requiring that K be a fixed, non-scaling, variable.

Finally, we can establish the final relationships between the full riemannian metrics $g_{ij}(t)$ and $g'_{ij} = g_{ij}(t+\delta t)$ on the two manifestly harmonically related slices t and $t+\delta t$. As in (95),

$$\partial_t g_{ij} = \partial_t(\psi^4 \gamma_{ij}) = g_{ik} g_{jl} \left[-2N(A^{kl} + \frac{1}{3} g^{kl} K) + (\nabla^k \beta^l + \nabla^l \beta^k) \right] . \quad (110)$$

Working out (110) gives

$$\begin{aligned} \partial_t g_{ij} &= \psi^4 \left[u_{ij} + \gamma_{ij} \partial_t (\psi \log \psi) \right] \\ &= V_{ij} + g_{ij} \partial_t (\psi \log \psi) , \end{aligned} \quad (111)$$

where

$$\begin{aligned} \partial_t(\psi \log \psi) &= \frac{2}{3} \left(D_k \beta^k + 6 \beta^k \partial_k \log \psi - \alpha K \psi^6 \right) \\ &= \partial_t (g/\gamma)^{1/2} = \partial_t (g)^{1/3} = \frac{2}{3} \left(\nabla_k \beta^k - NK \right) . \end{aligned} \quad (112)$$

Hence, $\partial_t \psi$ and $\partial_t g_{ij}$ are fully determined and we note that the no-scaling rules for β^k and K were essential.

Acknowledgments

This paper is based on a lecture given by James W. York Jr. at the 2nd Samos meeting, 1 September, 1998.

The authors thank the supporters of the 2nd Samos Meeting and its hosts, especially Spiros Cotsakis, for their warm hospitality and the beautiful setting of the meeting. They thank Mark Hannam and Sarah and Mark Rupright for help in preparing the manuscript. AA and JWY acknowledge support from National Science Foundation grant PHY-9413207.

References

1. Wheeler J.A. (1964) Geometrodynamics and the issue of the final state. In: DeWitt C., DeWitt B. (Eds.) Relativity, Groups, and Topology. Gordon and Breach, New York, 317-520.
2. Arnowitt R., Deser S., Misner C.W. (1962) The dynamics of general relativity. In: Witten L. (Ed.) Gravitation. Wiley, New York, 227-265.
3. Dirac P.A.M. (1958) The theory of gravitation in Hamiltonian form. Proc Roy Soc A246:333-343.
4. Dirac P.A.M. (1959) Fixation of coordinates in the Hamiltonian theory of gravitation. Phys Rev 114:924-930.
5. Lichnerowicz A. (1939) Problémes Globaux en Mecanique Relativiste. Hermann, Paris.

6. Choquet (Fourès)-Bruhat Y. (1952) Théorèm d'existence pour certains systèmes d'equations aux dérivées partielles non linéaires. Acta Math 88:141-225.
7. Choquet (Fourés)-Bruhat Y. (1956) Sur l'Integration des équations de la relativité générale. J Rat Mechanics and Anal 5:951-966.
8. Lichnerowicz A. (1944) L'intégration des équations de la gravitation relativiste et le problème des n corps. J Math Pures et Appl 23:37-63.
9. Choquet-Bruhat Y., York J.W. (1980) The Cauchy Problem. In: Held A. (Ed.) General Relativity and Gravitation, I. Plenum, New York, 99-172.
10. Anderson A. and York J.W. (1999) Fixing Einstein's equations. Phys Rev Lett 81:4384-4387.
11. Abrahams A., Anderson A., Choquet-Bruhat Y., York J.W. (1995) Einstein and Yang-Mills theories in hyperbolic form without gauge fixing. Phys Rev Lett 75:3377-3381.
12. Abrahams A., Anderson A., Choquet-Bruhat Y., York J.W. (1996) A non-strictly hyperbolic system for the Einstein equations with arbitrary lapse and shift. C.R. Acad Sci Paris Série IIb 323:835-841.
13. Abrahams A.,York J.W. (1997) 3+1 general relativity in hyperbolic form. In: Marck J-A., Lasota J-P. (Eds.) Relativistic Astrophysics and Gravitational Radiation. North Holland, Amsterdam, 179-190.
14. Abrahams A., Anderson A., Choquet-Bruhat Y., York J.W. (1997) Geometrical hyperbolic systems for general relativity and gauge theories. Class Quantum Grav 14:A9-A22.
15. Abrahams A., Anderson A., Choquet-Bruhat Y., York J.W. (1998) Hyperbolic formulation of general relativity. In: Olinto A.V., Frieman J.A., Schramm D.N. (Eds.) Proc 1996 Texas Symposium on Relativistic Astrophysics. World Scientific, Singapore, 601-603.
16. Bona C., Massó J. (1992) Hyperbolic evolution system for numerical relativity. Phys Rev Lett 68:1097-1099.
17. Bona C., Massó J., Seidel E., Stela J. (1995) New formalism for numerical relativity. Phys Rev Lett 75:600-603.
18. Choquet-Bruhat Y., Ruggeri T. (1983) Hyperbolicity of the 3+1 system of Einstein equations. Commun Math Phys 89:269-275.
19. Choquet-Bruhat Y., York J.W. (1995) Geometrical well posed systems for the Einstein equations. C.R. Acad Sci Paris, Série I t321:1089-1095.
20. Choquet-Bruhat Y., York J.W. (1996) Mixed Elliptic and Hyperbolic Systems for the Einstein Equations. In: Ferrarese G. (Ed.) Gravitation, Electromagnetism and Geometric Structures. Pythagora Editrice, Bologna, Italy, 55-74.
21. Choquet-Bruhat Y., York J.W. (1997) Well posed reduced systems for the Einstein equations. Banach Center Publications, Part 1 41:119-131.
22. Choquet-Bruhat Y., York J.W., Anderson A. (1998) Curvature-based hyperbolic systems for general relativity. To appear in: Piran T. (Ed.) Proc 1997 Marcel Grossmann Meeting, gr-qc/9802027.
23. Fischer A., Marsden J. The Einstein evolution equations as a first-order quasi-linear symmetric hyperbolic system. I. Commun Math Phys 28:1-38.
24. Friedrich H. (1985) On the hyperbolicity of Einstein and other gauge field-equations. Commun Math Phys 100:525-543.
25. Friedrich H. (1996) Hyperbolic reductions for Einstein's equations. Class Quantum Grav 13:1451-1459.

26. Frittelli S., Reula O. (1994) On the Newtonian limit of general relativity. Commun Math Phys 166:221-235.
27. Frittelli S., Reula O. (1996) First-order symmetric hyperbolic Einstein equations with arbitrary gauge. Phys Rev Lett 76:4667-4670.
28. van Putten M.H.P.M., Eardley D.M. (1996) Nonlinear wave equations for relativity. Phys Rev D53:3056-3063.
29. Anderson A. and York J.W. (1998) Hamiltonian time evolution for general relativity. Phys Rev Lett 81:1154-1157.
30. York J.W. (1979) Kinematics and dynamics of general relativity. In: Smarr L. (Ed.) Sources of Gravitational Radiation, Cambridge Univ Press, Cambridge, 83-126.
31. Teitelboim C. (1982) Quantum mechanics of the gravitational field. Phys Rev D25:3159-3179.
32. Ashtekar A. (1988) New Perspectives in Canonical Gravity. Bibliopolis, Naples.
33. Ashtekar A. (1987) New Hamiltonian formulation of general relativity. Phys Rev D36:1587-1602.
34. York J.W. (1973) Conformally invariant orthogonal decomposition of symmetric tensors on riemannian manifolds and the initial-value problem of general relativity. J Math Phys 14:456-464.
35. O'Murchadha N., York J.W. (1974) Initial-value problem of general relativity. I. General formulation and physical interpretation. Phys Rev D10:428-436.
36. Teitelboim C. (1973) How commutators of constraints reflect the spacetime structure. Ann Phys (NY) 79:542-557.
37. Teitelboim C. (1977) Supergravity and square root of constraints. Phys Rev Lett 38:1106-1110.
38. Friedrich H. (1991) On the global existence and the asymptotic-behaviour of solutions to the Einstein-Maxwell-Yang-Mills equations. J Diff Geom 34:275-345.
39. Choquet-Bruhat Y., Noutchegueme N. (1986) Système hyperbolique pour les équations d'Einstein avec sources. C.R. Acad Sc Paris, Série I t303:259-263.
40. Leray J. (1952) Hyperbolic Differential Equations. I.A.S lecture notes, Princeton.
41. Leray J., Ohya Y. (1967) Équations et systèmes non-linéaries, hyperboliques nonstricts. Math Ann 170:167-205.
42. Anderson A., Abrahams A., Lea C. (1998) Curvature-based gauge-invariant perturbation theory for gravity: a new paradigm. Phys Rev D5806:4015.
43. Choquet-Bruhat Y. (1997) High frequency oscillations of Einstein geometry. In: Ibragimov N., Mahomed F. (Eds.) Modern Group Analysis. World Scientific, Singapore, 17-34.
44. Choquet-Bruhat Y., Greco A. (1997) Interactions of gravitational and fluid waves. Circ Mat di Palermo 38:112-121.
45. O'Murchada N., York J.W. (1973) Existence and uniqueness of solutions of the Hamiltonian constraint of general relativity on compact manifolds. J Math Phys 14:1551-1557.
46. O'Murchadha N., York J.W. (1974) Initial-value problem of general relativity. II. Stability of solutions of the initial-value equations. Phys Rev D10:437-446.
47. Isenberg J., Moncrief V. (1994) Constraint equations with non-constant mean curvature. In: Flato M., Kerner R., Lichnerowicz A. (Eds.) Physics on Manifolds. Kluwer, Dordrecht, The Netherlands, 295-301.

48. York J.W. (1974) Covariant decompositions of symmetric tensors in the theory of gravitation. Ann Inst Henri Poincaré, Section A 21:319-332.
49. York J.W. (1999) Conformal "thin-sandwich" data for the initial-value problem of general relativity. Phys Rev Lett 82:1350-1353.
50. York J.W. (1972) Role of conformal three-geometry in the dynamics of gravitation. Phys Rev Lett 28:1082-1085.
51. Anderson A., Choquet-Bruhat Y., York J.W. (1997) Einstein-Bianchi hyperbolic system for general relativity. Topol Meth Nonlinear Anal 10:353-373.
52. Bel L. (1958) Définition d'une densité d'énergie et d'un état de radiation totale. C.R. Acad Sci Paris 246:3105-3108.
53. Bel L. (1961) L' identitie de Bianchi. Thése, University of Paris, Paris.

Generalized Bowen–York Initial Data

R. Beig

Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5,
A–1090 Wien, Austria

Abstract. A class of vacuum initial-data sets is described which are based on certain expressions for the extrinsic curvature first studied and employed by Bowen and York. These expressions play a role for the momentum constraint of General Relativity which is analogous to the role played by the Coulomb solution for the Gauß-law constraint of electromagnetism.

1 Introduction

In this lecture I wish to study a specific class of solutions to the initial-value constraints in vacuo, or, rather the 'momentum' part of these equations, which are a generalization of ones first put forward by Bowen and York ([1],[2]). While these solutions are certainly special, they turn out to be very useful. In particular, many of the initial-data sets currently used by numerical relativists are Bowen–York initial data (BY initial data), in the sense that they are based on the explicit extrinsic curvature expressions first written down in ([1],[2]). Although we shall in this work mainly be concerned with the momentum constraints, the solutions we shall study can only be understood as ingredients to solutions to the full set of initial-value constraints.

We first recall the notion of an initial-data set (IDS). This consists of a triple $(\bar{\Sigma}, \bar{h}_{ij}, \bar{K}_{ij})$, where $\bar{\Sigma}$ is a 3-manifold, $\bar{h}_{ij}$ a positive-definite metric on $\bar{\Sigma}$ and $\bar{K}_{ij}$ a symmetric tensor field on $\bar{\Sigma}$. This IDS is called a **vacuum IDS**, if the following system of equations is satisfied

$$\bar{D}^j(\bar{K}_{ij} - \bar{h}_{ij}\bar{K}) \quad = 0 \tag{1}$$

$$\bar{\mathcal{R}} + \bar{K}^2 - \bar{K}^{ij}\bar{K}_{ij} = 0, \tag{2}$$

where $\bar{D}_i$ is the Levi Civita covariant derivative associated with $\bar{h}_{ij}$, $\bar{\mathcal{R}}$ the scalar curvature and $\bar{K} := \bar{h}^{ij}\bar{K}_{ij}$. Given a vacuum IDS, there is a spacetime M with Ricci flat Lorentz metric $g_{\mu\nu}$, in which $\bar{\Sigma}$ is a Cauchy surface and $\bar{h}_{ij}$ (resp. $\bar{K}_{ij}$) are the intrinsic metric (resp. extrinsic curvature) induced on $\bar{\Sigma}$.

We now review the 'conformal method' for solving the vacuum initial-value constraints. This yields solutions $(\bar{h}_{ij}, \bar{K}_{ij})$ of Equ.'s (1,2) for which $\bar{K} \equiv$ const. Then, if we write

$$\bar{K}_{ij} = \bar{\kappa}_{ij} + \frac{1}{3}\bar{h}_{ij}\bar{K}, \tag{3}$$

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 55–69, 2000.

it follows that

$$\bar{D}^j \bar{\kappa}_{ij} = 0 \qquad \text{and} \qquad \bar{\kappa} = \bar{\kappa}_{ij} \bar{h}^{ij} = 0. \tag{4}$$

Thus $\bar{\kappa}_{ij}$ is a "transverse-tracefree" (TT-)tensor. The conformal method rests on two identities.

Fact a): Let h_{ij} and $\bar{h}_{ij}$ be two conformally related metrics, i.e.

$$\bar{h}_{ij} = \phi^4 h_{ij}, \qquad \phi > 0. \tag{5}$$

Then, defining the conformal Laplacian L_h to be

$$L_h := -h^{ij} D_i D_j + \frac{1}{8}\mathcal{R}[h] = -\Delta + \frac{1}{8}\mathcal{R} \tag{6}$$

we have that

$$L_{\bar{h}}(\phi^{-1}\psi) = \phi^{-5} L_h \psi, \qquad \phi > 0. \tag{7}$$

Setting in (7) $\phi = \psi$, it follows that

$$L_{\bar{h}} 1 = \frac{1}{8}\mathcal{R}[\bar{h}] = \phi^{-5} L_h \phi. \tag{8}$$

Fact b): Suppose $K_{ij} = K_{(ij)}$ is trace-free. Then

$$\bar{D}^j \bar{K}_{ij} = \phi^{-6} D^j K_{ij}, \qquad \phi > 0, \tag{9}$$

where $\bar{K}_{ij} = \phi^{-2} K_{ij}$. Combining these facts we can make the following observation: Suppose K_{ij} is a TT-tensor with respect to the metric h_{ij}. Then, for any $\phi > 0$ and any constant $\bar{K}$,

$$\bar{K}_{ij} = \phi^{-2} K_{ij} + \frac{\bar{K}}{3}\phi^2 h_{ij} \tag{10}$$

satisfies Equ. (1). Furthermore, if ϕ satisfies

$$L_h \phi = \frac{1}{8} K_{ij} K^{ij} \phi^{-7} - \frac{1}{12}\bar{K}^2 \phi^5, \tag{11}$$

then $\bar{h}_{ij} = \phi^4 h_{ij}$ satisfies Equ. (2). Thus, solving the constraints (1,2) amounts to choosing a 'background metric' h_{ij}, finding a TT-tensor K_{ij} with respect to h_{ij} and solving Equ. (11) for given (h_{ij}, K_{ij}) and a choice of constant $\bar{K}$.

We will need global conditions in order for the above program to go through. The two cases of greatest interest is, firstly, the case where $\bar{\Sigma}$ is compact ("cosmological case"), and here it is natural to take the background fields (h_{ij}, K_{ij}) to be defined on $\Sigma = \bar{\Sigma}$. Depending on the global conformal nature of h_{ij}, the sign of $\bar{K}$ and on whether $\bar{\kappa}_{ij}$ is zero or non-zero, there is an exhaustive list of possibilities [3] for which Equ. (11) can be solved.

The other case is that where $(\bar{\Sigma}, \bar{h}_{ij}, \bar{K}_{ij})$ is asymptotically flat. This means that there is a compact subset $K \subset \bar{\Sigma}$, so that $\bar{\Sigma} \setminus K$ consists of a finite number of asymptotic ends. An asymptotic end is a set diffeomorphic to $\mathbf{R}^3 \setminus B$, where B is a closed ball. Furthermore, in the coordinate chart given by this diffeomorphism, $\bar{h}_{ij}$ should satisfy

$$\bar{h}_{ij} - \delta_{ij} = O\left(\frac{1}{\bar{r}}\right), \qquad \bar{r}^2 = \bar{x}^i \bar{x}^j \delta_{ij}, \tag{12}$$

$$\bar{K}_{ij} = O\left(\frac{1}{\bar{r}^2}\right) \tag{13}$$

and $\partial \bar{h}_{ij} = o(1/\bar{r}^2), \ldots, \partial \bar{K}_{ij} = O(1/\bar{r}^3), \ldots$ for a few derivatives. Since we require $\bar{K} = \text{const}$, it follows that $\bar{K} = 0$.

Now the problem of solving the constraints takes the following form: One first picks an asymptotically flat 3-metric h_{ij} on $\bar{\Sigma}$ and a TT-tensor K_{ij} on $(\bar{\Sigma}, h_{ij})$. One then has to solve Equ. (11) with $\bar{K} = 0$ and the boundary condition that $\phi \to 1$ at infinity.

A very convenient alternative is to take (h_{ij}, K_{ij}) to be defined on a compact manifold Σ, the many-point compactification of $\bar{\Sigma}$. The "infinities" of $\bar{\Sigma}$ are now replaced by a finite number of points $\Lambda_\alpha \in \Sigma$, which we call punctures. The Equ. (11) should then be replaced by

$$L_h \phi = \frac{1}{8} K_{ij} K^{ij} \phi^{-7} + 4\pi (c_1 \delta_1 + c_2 \delta_2 + \ldots), \tag{14}$$

where δ_α is the delta distribution supported at Λ_α and c_α are positive constants. Note that $\bar{K} = 0$. Let h_{ij} be a smooth metric on Σ and K_{ij} be smooth on Σ except, perhaps, at the points Σ_α, where it may blow up like $1/r^4$, where $r^2 = \delta_{ij} x^i x^j$ with x^j being Riemann normal coordinates centered at Λ_α. Then a solution ϕ of Equ. (14) behaves like $1/r$ near Λ_α. If a positive global solution of Equ. (14) exists, then $\bar{h}_{ij} = \phi^4 h_{ij}$, $\bar{K}_{ij} = \phi^{-2} K_{ij}$ satisfy the conditions of asymptotic flatness in the "inverted" coordinates $\bar{x}^i = x^i / r^2$. For the conditions under which $\phi > 0$ exists, see e.g. [4].

2 The Gauß law constraint of electromagnetism

Before turning to the methods for obtaining TT-tensors, it is very instructive to use, by means of analogy, the equation

$$\operatorname{div} E = D^i E_i = 0 \tag{15}$$

on a Riemann 3-manifold (Σ, h_{ij}), that-is-to-say the Gauß constraint of electrodynamics. This, like the TT-condition, is an **underdetermined elliptic system**. This means that the "symbol map", namely the linear map $\sigma(k)$, sending

$$e_i \in \mathbf{R}^3 \mapsto k^i e_i \in \mathbf{R} \qquad (k \in \mathbf{R}^3, k \neq 0) \tag{16}$$

is onto. (The symbol map is essentially the Fourier transform of the highest-derivative term of a partial differential operator.) Note, first of all, that, when $\bar{h}_{ij} = \phi^4 h_{ij}$ $(\phi > 0)$, $\bar{E}_i = \phi^{-2} E_i$, we have that

$$\bar{D}^i \bar{E}_i = \phi^{-6} D^i E_i. \tag{17}$$

We will thus again take Σ to be compact, imagining that we are either in the cosmological case or that Σ is the conformally compactified, asymptotically flat 3-space $\bar{\Sigma}$. Our first point is to show two different ways of solving (15). To describe the first one, which is the analogue of the "York method" for solving the momentum constraint, we observe that the gradient operator grad, sending C^∞-functions on Σ to $\Lambda^1(\Sigma)$, the 1-forms on Σ, is (minus) the formal adjoint of div under the natural L^2-inner product on (Σ, h_{ij}). It is thus, on general grounds (see Appendix on Sobolev Spaces and Elliptic Operators in [5]) true that there is a direct-sum decomposition of $\Lambda^1(\Sigma)$ as

$$\Lambda^1(\Sigma) = \text{grad } (C^\infty(\Sigma)) \oplus \ker \text{ div}, \tag{18}$$

and this decomposition is orthogonal in the L^2-sense. Furthermore the second summand in (18) is infinite-dimensional. The relation (18) tells us how to find the elements of ker div, i.e. the solutions of Equ. (15). Namely, pick an arbitrary 1-form ω_i and write

$$\omega_i = D_i \varphi + E_i, \tag{19}$$

where $E_i \in \ker$ div. Hence

$$D^i \omega_i = \text{div grad } \varphi = \Delta \varphi. \tag{20}$$

Since the left-hand side of (20) is a divergence, it is orthogonal to $\ker \Delta$. Thus Δ^{-1} div ω exists. Consequently, E_i can be written as

$$E = [\mathbf{1} - \text{grad (div grad)}^{-1} \text{ div}]\omega, \tag{21}$$

and this is the general solution of (15). The Equ. (21) is a special case of the way to solve an underdetermined elliptic system. It has the feature that it is non-local, since it involves the operation of taking the inverse of div grad.

There is another way of solving Equ. (15). Observe that any 1-form E_i of the form (ε_{ijk} is the volume element on Σ)

$$E_i = \varepsilon_i{}^{jk} D_j \mu_k, \tag{22}$$

or

$$E = \text{rot } \mu$$

solves Equ. (15). Is this the general solution? The answer is: yes, up to at most a finite-dimensional space, namely the **harmonic** 1-forms, i.e. the elements of $\Lambda^1(\Sigma)$ which are annihilated by $\Delta_H = (\text{rot})^2 - \text{grad div}$. More specifically we have

$$\ker\ \text{div} = \ker \Delta_H \oplus\ \text{rot}\ (\Lambda^1(\Sigma)). \tag{23}$$

Note that $\ker \Delta_H = \ker\ \text{rot}\ \cap \ker\ \text{div}$. The relation (23) gives a refinement of the decomposition (20). Stated in a more fancy way, the decomposition (23) expresses the fact that the de Rham cohomology group $H^2 = \ker\ \text{div}/\text{rot}\ \Lambda^1(\Sigma)$ is isomorphic to $\ker \Delta_H$. Let us for simplicity assume that $\ker \Delta_H$ is trivial. In Wheeler's words [6] we assume that there is no "charge without charge". This will for example be the case when (Σ, h_{ij}) is the standard 3-sphere. More generally, this is the case when (Σ, h_{ij}) is of constant positive curvature (Exercise: prove this). This implies the following: Let E_i be an arbitrary solution of Equ. (15), and $S \subset \Sigma$ be an arbitrary embedded 2-sphere. Then the integral $\int_S E_i dS^i$ is zero, i.e.

$$\int_S E_i dS^i = 0. \tag{24}$$

Suppose now that Σ contains some punctures, at which we are willing to allow $D^i E_i$ to become singular. Then Equ. (24) will in general no longer be valid, and $E =\ \text{rot}\ \mu$ will no longer be the general solution to Equ. (15). Suppose for simplicity that Σ is diffeomorphic to the 3-sphere. If there is just one puncture Λ_1 (not to be confused with $\Lambda^1(\Sigma)$!), then (24) is clearly still valid, whether S endores Λ_1 or it doesn't. But suppose there are two punctures, Λ_1 and Λ_2. Then, either S does not enclose either of them in which case (24) is valid or it does, in which case (24) can not be expected to hold in general. The value

$$Q = \int_S E_i dS^i, \tag{25}$$

in the latter case, constitutes a 1-parameter class of obstructions to the existence of μ_i such that $E =\ \text{rot}\ \mu$. (Of course, Q depends only on the homology class of S, not S itself.) It is thus desirable to split the general E with div $E = 0$ as a sum of one of the form $E =\ \text{rot}\ \mu$ and a set of fields parametrized by Q. Imagine this latter set of fields to be distributions which are orthogonal to fields in the first summand. Thus these will be of the form $E =\ \text{grad}\ \varphi$, and

$$\text{div grad}\ \varphi = \rho, \tag{26}$$

where the distribution ρ is supported on $\Lambda_1 \cup \Lambda_2$ and has to satisfy

$$\int_\Sigma \rho dV = 0. \tag{27}$$

Thus we set

$$\rho = Q(\delta_1 - \delta_2), \tag{28}$$

where δ is the Dirac delta distribution. If φ is the solution (unique up to addition of a constant) of Equ. (27), $E = \text{grad } \varphi$ and the normal to S points towards Λ_2, then Equ. (25) is fulfilled.

Let us be more specific. When (Σ, h_{ij}) is the standard 3-sphere, it can be imagined to be the 1-point conformal compactification (conformal compactification by inverse stereographic projection from the origin, say) of flat $\mathbf{R}^3 = \bar{\Sigma}$ (which in turn can be viewed to be a standard $t = \text{const}$ hyperplane of the Minkowski spacetime of Special Relativity). Suppose the origin of $\bar{\Sigma}$ corresponds to Λ_1 and Λ_2 is the antipode of Λ_1 on S^3. Thus Λ_1 should be viewed as a point where the field becomes singular and Λ_2 as the point-at-infinity. (In the gravity case, of course, singularities do not make good sense, sp all punctures will play the role of points-at-infinity. That this is possible, is due to the fact that $\bar{h}_{ij}$ on $\bar{\Sigma}$ is not a fixed element of the theory, but $\bar{h}_{ij} = \phi^4 h_{ij}$, and ϕ satisfies Equ. (14).) We now undo the stereographic projection and write

$$\bar{E}_i = \phi^{-2} D_i \varphi, \tag{29}$$

where φ solves (26) and ϕ solves

$$L_h \phi = 4\pi \delta_1. \tag{30}$$

The resulting field $\bar{E}_i$ is nothing but the Coulomb field on $\bar{\Sigma} = \mathbf{R}^3$ with charge Q sitting at the origin.

The above situation can be slightly generalized. Let ρ be an arbitrary distribution on Σ such that

$$\int_\Sigma \rho dV = 0. \tag{31}$$

In particular ρ could be a smooth function. Then, if there are no harmonic 1-forms, the general solution of

$$D^i E_i = \rho \tag{32}$$

is of the form

$$E = \text{rot } \mu + \text{grad } \varphi, \tag{33}$$

where

$$\text{div grad } \varphi = \rho. \tag{34}$$

The second term in Equ. (33) could be called the generalized Coulomb field corresponding to the source ρ or, alternatively, the “longitudinal” solution of Equ. (32).

3 The momentum constraint

We will now turn to the gravity case, i.e. the momentum constraints. The solutions originally due to Bowen and York (resp. our generalization thereof) will turn out to be close analogues of the Coulomb field (resp. the generalized Coulomb field), as described above. When (Σ, h_{ij}) is again taken to be standard S^3 with antipodal punctures Λ_1, Λ_2 there will be a 10-parameter set of sources for the TT-condition $D^j K_{ij} = 0$, $K = 0$, so that, under stereographic projection relative to Λ_1, the TT-tensor on punctured $\mathbf{R}^3$ corresponding to the longitudinal solution of the inhomogeneous TT-condition on Σ contain exactly the ones written down by Bowen and York. How does the number 10 enter here? In the Maxwell case we had 1, which was the null space of grad, which in turn is the adjoint of the operator div. In the gravity case 10 arises as the null space of the adjoint to the operator div, acting on symmetric, trace-free tensors, namely the conformal Killing operator. This null space, in turn, is the space of conformal Killing vectors, which has at most 10 dimensions on a 3-manifold Σ. We now have to explain these things in more detail. We again take (Σ, h_{ij}) to be compact. The underdetermined elliptic system

$$(\text{div } K)_i = D^j K_{ij} = 0, \qquad K = 0 \tag{35}$$

is, in complete analogy with the Maxwell case, solved by the "York decomposition"

$$Q_{ij} = (LW)_{ij} + K_{ij}, \tag{36}$$

where Q_{ij} is an arbitrary symmetric, trace-free tensor on (Σ, h_{ij}) and L is the conformal Killing operator

$$(LW)_{ij} := D_i W_j + D_j W_i - \frac{1}{3} h_{ij} D^\ell W_\ell, \tag{37}$$

(not to be confused with the conformal Laplacian L_h) which is (-1/2 times) the adjoint of div. Hence, in order for K_{ij} to be in the null space of div, we have that

$$\text{div } \circ LW = \text{div } Q. \tag{38}$$

The operator on the left in Equ. (38) is elliptic with null space consisting of conformal Killing vectors (Proof: Contract with a covector λ and integrate by parts!). Thus, in complete analogy to Equ. (21) in the electromagnetic case (15), the general solution K_{ij} of the TT-condition, Equ. (35), is given by

$$K = [\mathbf{1} - L(\text{div } \circ L)^{-1} \text{div}]Q, \tag{39}$$

where Q_{ij} is an arbitrary symmetric, trace-free tensor. The above procedure works for an arbitrary (Σ, h_{ij}) with Σ compact. It would also work if Σ was asymptotically flat (see e.g. [7]). We now again ask the question as to whether there is a more explicit method for finding TT-tensors where (35) is solved "by differentiation" rather than "by integration", as in Equ. (39). There is a positive answer, but only in the case where (Σ, h_{ij}) is (locally) conformally flat. Luckily, this comprises many of the IDS's which are currently in use by the numerical relativists, as we shall describe in Sect. 5. The necessary and sufficient condition [8] for h_{ij} to be conformally flat is that the Cotton–York tensor $\mathcal{H}_{ij}$ defined by

$$\mathcal{H}_{ij} = \varepsilon_{k\ell(i} D^k \mathcal{R}^\ell{}_{j)} \tag{40}$$

be zero. Note that $\mathcal{H}_{ij}$ is always symmetric, trace-free. Also, as a consequence of the Bianchi identities, it is divergence-free. Thus $\mathcal{H}_{ij}$ is a TT-tensor with respect to h_{ij}. If h_{ij} is such that $\mathcal{H}_{ij}$ vanishes, it follows that the operator H_{ij}, obtained by linearization of $\mathcal{H}_{ij}$ at h_{ij}, is a (third-order) partial differential operator mapping symmetric tensors ℓ_{ij} (we take them also to be trace-free) into tensors which are TT with respect to h_{ij}. We refrain here from writing down the operator H_{ij} explicitly (see [9]). There now arises the question of whether K_{ij} given by $K_{ij} = H_{ij}(\ell)$ is the general TT-tensor. The answer, given in [9], is again "yes" up to a finite dimensional set of "harmonic TT-tensors", i.e. symmetric, trace-free tensors ℓ_{ij} which satisfy both $D^j \ell_{ij} = 0$ and $H_{ij}(\ell) = 0$. What is the condition for the absence of such harmonic TT-tensors? In allusion to Wheeler [6] we might describe this situation by the absence of "momentum without momentum". It is shown in [9], that this condition is exactly that, for any 2-surface S embedded in Σ and any conformal Killing vector (CKV) ξ^i on (Σ, h_{ij}) which is defined near S, there holds

$$\int_S K_{ij} \xi^i dS^j = 0 \tag{41}$$

for all TT-tensors K_{ij}. We now come to the inhomogeneous equation

$$D^\ell K_{i\ell} = j_i \tag{42}$$

which is the analogue of the Maxwell equation (32). It would be nice to have a very clear geometrical motivation for our expression for j_i, but we have to leave that for future work. The proposal is that j_i depends on a "charge density ρ" and a CKV η^i, as follows:

$$j_i(\eta) = -D^\ell(D_{[\ell}\eta_{i]}\rho) + \frac{2}{3}(D_i D_\ell \eta^\ell)\rho + \frac{2}{3} D_i D_\ell(\eta^\ell \rho) + \frac{2}{9} D_i((D_\ell \eta^\ell)\rho) + 4L_{i\ell}\eta^\ell \rho, \tag{43}$$

where $L_{ij} = \mathcal{R}_{ij} - \frac{1}{4} h_{ij}\mathcal{R}$. It is important that Equ. (42) behaves naturally under conformal rescalings of the metric, i.e. $\bar{h}_{ij} = \phi^4 h_{ij}$. This is the case since one can show that

$$\bar{j}_i(\eta) = \omega^{-6} j_i(\eta). \tag{44}$$

Thus, with $\bar{K}_{ij} = \omega^{-2} K_{ij}$, $\bar{K}_{ij}$ satisfies $\bar{D}^\ell \bar{K}_{i\ell} = \bar{j}_i$ provided K_{ij} solves Equ. (42).

Suppose we have an open region $\Omega \subset \Sigma$, bounded by S. Then it follows from (43) that

$$\int_S K_{ij} \lambda^i dS^j = \int_\Omega X(\lambda, \eta) \rho dV, \tag{45}$$

where

$$X(\lambda, \eta) = D^{[i}\lambda^{j]}D_{[i}\eta_{j]} + \frac{2}{3}[\lambda^i D_i D_j \eta^j + \eta^i D_i D_j \lambda^j] - \frac{2}{9}(D_i\lambda^i)(D_j\eta^j) + 4L_{ij}\lambda^i\eta^j. \tag{46}$$

The bilinear form X has an important geometrical meaning. Recall that the CKV's span a vector space $\mathbf{W}$ of at most ten dimensions. (The dimension ten is reached when Σ is simply connected, in which case the general CKV η^i can be characterized by the values of η^i at some point $p \in \Sigma$, together with those of $D_{[i}\eta_{j]}$, $D_j\eta^j$ and $D_iD_j\eta^j$, the so-called "conformal Killing data" [10].) The Lie commutator of vector fields on Σ induces a Lie algebra structure on $\mathbf{W}$. The form $X(\lambda, \eta)$, for η and λ **both** CKV's, is nothing but (1/3) times the Killing metric (see App. B of [11]). One can check by explicit computation that

$$X(\lambda, \eta) = \text{constant on } \Sigma, \tag{47}$$

when λ and η are both CKV's. Thus, from (45)

$$\int_S K_{ij}\lambda^i dS^j = X(\lambda, \eta)\int_\Omega \rho dV. \tag{48}$$

So, similar to the Maxwell case, ρ can not in general be a distribution concentrated at a single point. Rather, when $\Omega = \Sigma$ it follows that

$$X(\lambda, \eta)\int_\Sigma \rho dV = 0. \tag{49}$$

Equ. (50) is the necessary and sufficient condition in order for Equ. (42), with j_i given by (43), to have solutions. Note that, when dim $\mathbf{W}$ is 10, the Killing metric is non-degenerate. The solution of the equation

$$D^\ell K_{i\ell} = j_i(\eta) \tag{50}$$

becomes unique when we require K_{ij} to be longitudinal, i.e.

$$K_{ij} = (LW)_{ij}. \tag{51}$$

It remains to solve the elliptic equation

$$\text{div } \circ LW = j(\eta). \tag{52}$$

The most important case is again where (Σ, h_{ij}) is the standard three-sphere and

$$\rho = -2\pi(\delta_1 - \delta_2). \tag{53}$$

To write down explicitly the 10-parameter set of solutions, it is convenient to send Λ_2 to infinity by a stereographic projection. Then we have again Equ.

(49), but on $(\mathbf{R}^3, \delta_{ij})$, punctured at the origin. The CKV's on $\mathbf{R}^3$ fall into the following classes

$$ {}^1\eta^i(x) = Q^i, \qquad Q^i = \mathrm{const} \tag{54} $$
$$ {}^2\eta^i(x) = \varepsilon^i{}_{jk}S^j x^k, \qquad S^i = \mathrm{const} \tag{55} $$
$$ {}^3\eta^i(x) = Cx^i, \qquad C = \mathrm{const} \tag{56} $$
$$ {}^4\eta^i(x) = (x,x)P^i - 2(x,P)x^i, \qquad P^i = \mathrm{const}. \tag{57} $$

Here x^i are cartesian coordinates. We find

$$ {}^1K_{ij}(x) = \frac{3}{2r^2}[P_i n_j + P_j n_i - (\delta_{ij} - n_i n_j)(P,n)] \tag{58} $$
$$ {}^2K_{ij}(x) = \frac{6}{r^3}\varepsilon_{k\ell(i}S^k n^\ell n_{j)} \tag{59} $$
$$ {}^3K_{ij}(x) = \frac{C}{r^3}(3n_i n_j - \delta_{ij}) \tag{60} $$
$$ {}^4K_{ij}(x) = \frac{3}{2r^4}[-Q_i n_j - Q_j n_i - (\delta_{ij} - 5n_i n_j)(Q,n)]. \tag{61} $$

Here $n_i = x_i/r$. The constants P_i in (58) play the role of the linear ADM-momentum at $r = \infty$ and the S^i in (59) are the ADM angular momentum. Thus P_i and S_i are conserved under time evolution. If we had sent the point Λ_1 to infinity, or – what is the same – if we made an inversion of the form $\bar{x}^i = x^i/a^2r^2$ ("Kelvin transform"), the Q_i would be a^2 times the linear momentum at $r = 0$, viewed as another infinity. Similarly (54) goes over into (57) with $P^i = Q^i$ under Kelvin transform. The role of the constant C is less clear. It was used in [4], to construct IDS's which have future-trapped surfaces. The constant C does not correspond to a conserved quantity.

If K_{ij} was a sum of the expression in (58) and that in (59) the physical interpretation is that they characterize a single black hole with momentum P^i and spin S^i.

We now come to the technical result of this section. Suppose (Σ, h_{ij}) is of constant curvature, i.e.

$$ \mathcal{R}_{ijk\ell} = \frac{\mathcal{R}}{3}h_{k[i}h_{j]\ell}, \qquad \mathcal{R} = \mathrm{const}. \tag{62} $$

Suppose, further, we know a function (distribution) G satisfying

$$ \Delta G = \rho. \tag{63} $$

In the case when Σ is compact, this will exist provided that $\int_\Sigma \rho dV = 0$. Then there is an explicit expression for W_i solving Equ. (52), namely

$$ \begin{aligned} W_i &= \frac{1}{2}\eta^j D_j D_i G + \frac{3}{2}(D_i\eta_j)D^j G + (D_j\eta^j)D_i G + \frac{\mathcal{R}}{3}\eta_i G \\ &= -D^j[(D_{[j}\eta_{i]})G] - \frac{1}{6}D_i[(D_j\eta^j)G] + \frac{1}{2}D_i D_j(\eta^j G). \end{aligned} \tag{64} $$

For an outline of the proof, see [12]. It is easy to check that, when Σ is flat $\mathbf{R}^3$, $\rho = -2\pi\delta(x)$ and ξ^i runs through (54–57), the K_{ij}'s given by $(LW)_{ij}$ with W_i as in Equ. (64), reduce to (58–61).

4 Boosting a single black hole

In this section we present a simple perturbative calculation which should serve as a check whether the "boost-type" extrinsic curvature (58) gives a sensible result for the full IDS. We assume that (Σ, h_{ij}) is a standard three-sphere and we try to solve the Lichnerowicz equation (14) with Λ_1 and Λ_2 being south and north pole, respectively. The K_{ij} in (14) should be the one turning into (58) after stereographic projection. When K_{ij} vanishes and the three-sphere has radius $1/m$, the unique solution to (14) gives rise to the Schwarzschild solution of mass m. After stereographic projection the three-sphere punctured at Λ_1 and Λ_2 becomes flat $\mathbf{R}^3$, punctured at the origin. The Lichnerowicz conformal factor ϕ such that $\bar{h}_{ij} = \phi^4 \delta_{ij}$ for Schwarzschild is, as is well-known,

$$\phi = 1 + \frac{m}{2r}. \tag{65}$$

By the conformal invariance of the York procedure we can of course also start from flat, punctured $\mathbf{R}^3$. We are thus trying to solve

$$\Delta\phi = -\frac{1}{8} K_{ij} K^{ij} \phi^{-7}, \tag{66}$$

with K_{ij} given by Equ. (58), and ϕ should go to one at infinity and have a $1/r$-singularity near $r = 0$ so that $\phi^4 \delta_{ij}$ becomes asymptotically flat both near infinity and near $r = 0$. We have that

$$K_{ij} K^{ij} = \frac{9}{2r^4} [P_i P^i + 2(P_i n^i)^2]. \tag{67}$$

For ϕ we make the ansatz

$$\phi = 1 + \frac{m}{2r} + \psi, \tag{68}$$

where ψ should vanish at $r = \infty$ and be regular near $r = 0$. We only keep terms quadratic in P_i. It follows that

$$\Delta\psi = -\frac{9r^3}{16} \frac{P^2 + 2(P, n)^2}{(r + m/2)^7}. \tag{69}$$

Thus

$$\psi(x) = \frac{9}{4\pi \cdot 16} \int_{\mathbf{R}^3} \frac{r'^3}{|x - x'|} \frac{P^2 + 2(P, n')^2}{(r' + m/2)^7} dx'. \tag{70}$$

It is not difficult to find that

$$\lim_{x \to 0} \psi(x) = \frac{P^2}{8m^2} \tag{71}$$

and

$$\psi(x) = \frac{5P^2}{16mr} + O\left(\frac{1}{r^2}\right), \tag{72}$$

near $r = \infty$. While the constant m is only a formal parameter, the true "observables" are the ADM-energies at the two infinities. The ADM energy M at $r = \infty$ is given by

$$M = m + \frac{5P^2}{8m}. \tag{73}$$

The ADM energy $\bar{M}$ near $r = 0$ is obtained by noting that, near $r = 0$,

$$d\bar{s}^2 = \bar{h}_{ij}dx^i dx^j = \left[1 + \frac{P^2}{8m^2} + \frac{m}{2r} + O(r)\right]^4 \delta_{ij}dx^i dx^j. \tag{74}$$

After inversion $\bar{x}^i = (m/2)^2 x^i/r^2$, this results in

$$\begin{aligned} d\bar{s}^2 &= \left[1 + \frac{P^2}{8m^2} + \frac{2\bar{r}}{m} + O\left(\frac{1}{\bar{r}}\right)\right]^4 \left(\frac{m}{2}\right)^4 \frac{1}{\bar{r}^4}\delta_{ij}d\bar{x}^i d\bar{x}^j \\ &= \left[\frac{m}{2\bar{r}}\left(1 + \frac{P^2}{8m^2}\right) + 1 + O\left(\frac{1}{\bar{r}^2}\right)\right]\delta_{ij}d\bar{x}^i d\bar{x}^j. \end{aligned} \tag{75}$$

Consequently,

$$\bar{M} = m + \frac{P^2}{8m} + O(P^4). \tag{76}$$

Thus,

$$m = \bar{M} - \frac{P^2}{8\bar{M}} + O(P^4). \tag{77}$$

Inserting (77) into (73) we finally obtain

$$M = \bar{M} + \frac{P^2}{2\bar{M}} + O(P^4) \tag{78}$$

or

$$M^2 - P^2 = \bar{M}^2 + O(P^4). \tag{79}$$

Recall that with our ansatz for K_{ij} an observer near $r = \infty$ "sees" a hole at $r = 0$ with momentum P and one easily finds that an observer near $r = 0$ sees a hole at $r = \infty$ at momentum zero. Thus Equ. (79) says that the rest-masses at both asymptotic ends are equal. This equality expresses the "absence of gravitational radiation" on the spacetime slice Σ. If there was more than one hole, the analogue of Equ. (79) would contain for example potential energy-contributions from the mutual gravitational interaction between those holes, as in the famous calculation by Brill and Lindquist [13]. See [14] for interaction energies of more general geometries. Had we used as our ansatz for K_{ij} a sum of the one in Equ. (58) and the one in (61) we would obtain

$$M + \frac{\bar{P}^2}{2M} + O(\bar{P}^4) = \bar{M} + \frac{P^2}{2\bar{M}} + O(P^4), \tag{80}$$

where $\bar{P}$ is the ADM 3-momentum of the $r = 0$-end of Σ, which is related to Q by $\bar{P} = 4Q/M^2$. I thank U.Kiermayr for performing this computation.

Does, for the present IDS, the relation

$$M^2 - P^2 = \bar{M}^2 - \bar{P}^2 \tag{81}$$

hold exactly? In this connection we should point out that the present data is **not** the same as the one induced by the Schwarzschild spacetime of mass M on a boosted maximal slice, since the metric on such slices can not be conformally flat (see [15]).

5 More general initial-data sets with punctures

Let again (Σ, h_{ij}) be a compact, conformally flat manifold and η^i a CKV on Σ. Then consider the equation (50), namely

$$D^\ell K_{i\ell} = j_i(\eta), \tag{82}$$

with $j_i(\eta)$ given by Equ. (43). Although we assumed in our heuristic discussion that there are no "harmonic" TT-tensors on Σ, this is actually not required for (81) to make sense. The simplest case is where Σ is a standard three-sphere and ρ is a delta function concentrated at a finite number of points $\Lambda_\alpha \in \Sigma$ $(\alpha = 1, \ldots, N)$. We have to have

$$\int_\Sigma \rho dV = 0 \tag{83}$$

so that $N \geq 2$. The case $N = 2$ contains the situation discussed in the previous section. The case of general N is for some choices of η^i leads to the IDS's studied by Brandt and Brügmann [16].

Another interesting case is that where (Σ, h_{ij}) is $S^2 \times S^1(a)$, i.e. the unit-two sphere times the circle of length a. This is conformally flat but not of constant curvature. The equation

$$L_h \phi = 4\pi \delta_1 \tag{84}$$

has a unique positive solution ϕ (since $\mathcal{R} > 0$, see [17]). The manifold $(\bar{\Sigma} = \Sigma \setminus \Lambda_1, \bar{h}_{ij})$ with $\bar{h}_{ij} = \phi^4 h_{ij}$ is nothing but the (time-symmetric) Misner wormhole [18]. Taking two punctures Λ_1 and Λ_2 on Σ at the same location in S^2 and at opposite points in $S^1(a)$ and solving

$$L_h \phi = 4\pi(c_1 \delta_1 + c_2 \delta_2), \tag{85}$$

(which can be done by linearly superposing two solutions like that in (84)) we get for $(\bar{\Sigma}, \bar{h}_{ij})$ two asymptotically flat sheets joined by two Einstein–Rosen bridges [19]. These two time-symmetric IDS's can be turned into generalized Bowen–York ones by first solving

$$D^\ell K_{i\ell} = j_i(\eta) \tag{86}$$

for some CKV η^i and then solving

$$L_h\phi = \frac{1}{8}K_{ij}K^{ij}\phi^{-7} + 4\pi\sum_\alpha c_\alpha\delta_\alpha. \tag{87}$$

We just deal with Equ. (86) here. In the first case (1 puncture) we cannot have $\int \rho = 0$, so that η has to be in the null space of $X(\xi,\eta)$, where ξ runs through all CKV's on (Σ, h_{ij}). But, on this manifold, all conformal Killing vectors are in fact Killing vectors. (This would be true for any compact Riemannian manifold except standard S^3, see [20] or also App. A of [21].) Hence the CKV's just comprise rotations in the S^2-direction and a covariant constant vector in the S^1-direction. The latter, from Equ. (45), is allowed, but not the rotations. Thus, within the method of this paper, it is possible to boost a Misner wormhole into the direction connecting the two wormhole throats, but it is impossible to spin up a Misner wormhole. In the Einstein–Rosen case both options are available (see Bowen, York [22], Kulkarni et al. [23] and Bowen et al. [24]). It is not clear to me whether the solutions of (81) constructed by these authors, when viewed as ones on the handle manifold $S^2 \times S^1(a)$, are longitudinal or not.

Acknowledgement

This work was supported by Fonds zur Förderung der wissenschaftlichen Forschung in Österreich, Projekt Nr. P12626-PHY. I thank U.Kiermayr for a discussion which led to the discovery of an error in a previous version of this work.

References

1. Bowen J M 1979 Gen. Rel. Grav. **11**, 227
2. York J W 1989, in: Frontiers in Numerical Relativity, eds. C P Evans, L S Finn (Cambridge: Cambridge University Press)
3. Isenberg J 1995 Class. Quantum Grav. **12**, 2249
4. Beig R, Ó Murchadha N 1996 Class. Quantum Grav. **13**, 739
5. Besse A L 1987 Einstein Manifolds (Berlin Heidelberg New York: Springer)
6. Wheeler J A 1962 Geometrodynamics (New York: Academic Press)
7. Chaljub-Simon A 1981 Gen. Rel. Grav. **14**, 743
8. Schouten J 1921 Math. Z. **11**, 58
9. Beig R 1997 in: Mathematics of Gravitation, ed. P T Chrusciel, Banach Center Publ. **41**, 109

10. Geroch R 1969 Commun. Math. Phys. **13**, 180
11. Geroch R 1970 J. Math. Phys. **11**, 1955
12. Beig R 1997 ESI (Erwin Schrödinger Institute for Mathematical Physics) preprint 507
13. Brill D, Lindquist R 1963 Phys. Rev. **131**, 471
14. Giulini D 1996 Class. Quantum Grav. **7**, 1271
15. York J W 1980 in: Essays in General Relativity, ed. F J Tipler (New York: Academic Press)
16. Brandt S, Brügmann B 1997 Phys. Rev. Lett. **78**, 3606
17. Lee J M, Parker T H 1987 Bull. Am. Math. Soc. **17**, 37
18. Misner C W 1960 Phys. Rev. **118**, 1110
19. Misner C W 1963 Ann. Phys. **24**, 102
20. Obata M 1971 J. Diff. Geom. **6**, 237
21. Beig R, Ó Murchadha N 1996 Commun. Math. Phys. **176**, 723
22. Bowen J M, York J W 1980 Phys. Rev. **21**, 2047
23. Kulkarni A D, Shepley L C and York J W 1983 Phys. Lett. **86A**, 228
24. Bowen J M, Rauber J and York JW 1984 Class. Quantum Grav. **1**, 591

The Reduced Hamiltonian of General Relativity and the σ-Constant of Conformal Geometry

Arthur E. Fischer[1] and Vincent Moncrief[2]

[1] Department of Mathematics, University of California, Santa Cruz, CA 95064, USA. aef@cats.ucsc.edu

[2] Departments of Mathematics and Physics, Yale University, New Haven, CT 06511, USA. vincent.moncrief@yale.edu

Abstract. For the problem of the Hamiltonian reduction of Einstein's equations on a 3+1 vacuum spacetime that admits a foliation by constant mean curvature *(CMC)* compact spacelike hypersurfaces M that satisfy certain topological restrictions, we introduce a *dimensionless non-local time-dependent* reduced Hamiltonian system

$$H_{\text{reduced}} : \boldsymbol{R}^{-} \times P_{\text{reduced}} \longrightarrow \boldsymbol{R}$$

where the reduced Hamiltonian is given by

$$H_{\text{reduced}}(\tau, \gamma, p^{TT}) = -\tau^3 \int\limits_M \varphi^6 \mu_\gamma = -\tau^3 \int\limits_M \mu_g = -\tau^3 \mathrm{vol}(M, g) \quad .$$

For compact connected oriented 3-manifolds of Yamabe type -1, we establish the following properties for this reduced system:

1. $H_{\text{reduced}}(\tau, \gamma, p^{TT})$ is a monotonically decreasing function of t unless $p^{TT} = 0$ and $\gamma = \tilde{\gamma}$ is hyperbolic, at which point $H_{\text{reduced}}(\tau, \tilde{\gamma}, 0)$ is constant in time.
2. For $\tau \in \boldsymbol{R}^{-}$ fixed, $H_{\text{reduced}}(\tau, \gamma, p^{TT})$ has a unique (up to isometry) critical point at $(\tilde{\gamma}, 0)$ which is a strict local minimum (in the non-isometric directions).
3. For $\tau \in \boldsymbol{R}^{-}$ fixed, the σ-constant of M is related to H_{reduced} by

$$\sigma(M) = -\tfrac{2}{3}\Big(\inf_{(\gamma, p^{TT}) \in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, p^{TT}) \Big)^{2/3} \quad .$$

If M is a hyperbolic manifold, then we conjecture that $(\tilde{\gamma}, 0)$ is a strict global minimum of $H_{\text{reduced}}(\tau, \gamma, p^{TT})$ which, as part of our work, is equivalent to the conjecture that the σ-constant of M is realized by the unique hyperbolic geometry on M. If M is not a hyperbolic manifold, then the σ-constant is never realized by a metric on M but is only approached as a limit. In this case, the Einstein flow seeks to attain the σ-constant asymptotically insofar as the reduced Hamiltonian is monotonically seeking to decay to its infimum, although possible obstructions, such as the formation of black holes, may prevent any particular solution from approaching the σ-constant asymptotically. Further applications and developments in higher dimensions are discussed.

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 70–101, 2000.

1 Introduction, Notation, and Background

One of the goals of general relativity is to write Einstein's field equations as an *unconstrained* Hamiltonian dynamical system in which the constraint equations have been implicitly solved, the gauge variables have been specified, and the remaining variables are freely specifiable dynamical variables of the unconstrained system. This problem is known as *the problem of Hamiltonian reduction of Einstein's equations.*

In a recent series of papers, the authors ([13], [14], [15]) have resolved this problem of reduction for the Einstein vacuum field equations in $3+1$ dimensions in the case that the vacuum spacetime admits a foliation by constant mean curvature *(CMC)* compact spacelike hypersurfaces that satisfy certain topological restrictions. This resolution is a two-step procedure. The first step involves finding a suitable reduced phase space of unconstrained dynamical degrees of freedom. The second step involves finding a reduced Hamiltonian on this reduced phase space which is the true non-vanishing Hamiltonian of the theory.

We find that for *CMC* spacetimes with spacelike hypersurfaces M that satisfy certain conditions, $3+1$ dimensional reduction can be completed much as in the $2+1$ dimensional case ([29], [30]). In both cases, one gets as the reduced phase space the *cotangent bundle* $T^*\mathcal{T}_M$ of the *Teichmüller space*

$$\mathcal{T}_M = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}_0}$$

of conformal structures on M (see Sect. 2.1 for the definitions of the spaces involved) and one gets reduction of the full classical Hamiltonian system with constraints to a *non-local time-dependent reduced Hamiltonian system without constraints on the contact manifold* $\boldsymbol{R}^- \times T^*\mathcal{T}_M$, where $\boldsymbol{R}^- = (-\infty, 0)$.

In [15] we introduced the reduced phase space P_{reduced}, given by

$$\begin{aligned} P_{\text{reduced}} = \{(\gamma, p^{TT}) \mid \gamma \in \mathcal{M}_{-1}(\Sigma_p) \text{ and} \\ p^{TT} \text{ is transverse-traceless with respect to } \gamma\,\} \end{aligned}$$

where $\mathcal{M}_{-1}$ is the space of Riemannian metrics with scalar curvature -1 and p^{TT} is a 2-contravariant symmetric tensor density field on M transverse (divergence-free) and traceless with respect to γ. Associated with P_{reduced} is the *reduced contact manifold*

$$\boldsymbol{R}^- \times P_{\text{reduced}}$$

with *reduced contact variables* $(\tau, \gamma, p^{TT}) \in \boldsymbol{R}^- \times P_{\text{reduced}}$. In the case that the underlying 3-manifold M is of Yamabe type -1, described in Sect. 2.2, the cotangent bundle $T^*\mathcal{T}_M$ of the Teichmüller space $\mathcal{T}_M$ can be represented as

$$T^*\mathcal{T}_M \approx P_{\text{reduced}}/\mathcal{D}_0 \ .$$

We remark that, for clarity of exposition, in this paper we work on P_{reduced} rather than the fully reduced phase space $P_{\text{reduced}}/\mathcal{D}_0$. However, with a mild additional assumption, namely, that the *degree of symmetry* of the underlying manifold M is zero (i.e., that M not support an effective $SO(2)$-action), then the resulting fully reduced phase space is an infinite-dimensional manifold and our results in this paper transfer to this manifold. Without this additional assumption on M, our results still remain valid in a stratified sense.

For the reduced system, the time parameter $\tau = \text{tr}_g k$ is the trace of the extrinsic curvature and is the parameter of a family of monotonically increasing constant mean curvature compact spacelike hypersurfaces in a neighborhood of the given initial one. In [15], we chose $t = \frac{4}{3}\tau$ as a temporal coordinate condition fixing t in terms of τ. In this gauge, the reduced Hamiltonian becomes the spatial volume of the gravitational variables (g, π) on the $\tau =$ *constant* hypersurfaces, expressed in terms of the reduced canonical contact variables (τ, γ, p^{TT}) of these hypersurfaces, so that

$$H^{\text{old}}_{\text{reduced}}(\tau, \gamma, p^{TT}) = \int_M \varphi^6 \mu_\gamma = \int_M \mu_{\varphi^4\gamma} = \int_M \mu_g = \text{vol}(M, g) \ . \tag{1}$$

Thus, in the gauge $t = \frac{4}{3}\tau$, *the reduced Hamiltonian is the volume functional of the spatial hypersurfaces.*

In the current paper we make a different choice of time function that renormalizes the reduced Hamiltonian of [15] resulting in a renormalized Hamiltonian that is a dimensionless quantity. The advantages of such a renormalization are discussed below.

Thus, instead of taking the time function t to be proportional to the mean curvature τ, here we take

$$t = \frac{2}{3\tau^2} \tag{2}$$

as the temporal gauge condition. With this new time function, the new reduced Hamiltonian is given by

$$H^{\text{new}}_{\text{reduced}}(\tau, \gamma, p^{TT}) = -\tau^3 \int_M \varphi^6 \mu_\gamma = -\tau^3 \int_M \mu_{\varphi^4\gamma} = -\tau^3 \int_M \mu_g = -\tau^3 \text{vol}(M, g) \tag{3}$$

where the conformal factor $\varphi = \varphi(\tau, \gamma, p^{TT})$ is the unique positive solution of the Lichnerowicz equation (11) given below, and where $\tau \in \boldsymbol{R}^-$, the conformal metric $\gamma \in \mathcal{M}_{-1}$ satisfies $g = \varphi^4\gamma$, and p^{TT} is a transverse-traceless (with respect to γ) tensor density.

This new choice of time function is motivated by two factors. Firstly, with this choice of time function, the reduced Hamiltonian is dimensionless, and secondly, the range of the coordinate time t is recalibrated so that the big bang occurs at $t = 0$ instead of at $t = -\infty$.

That H_{reduced} is dimensionless follows at once from the fact that in *physical units*, τ has dimensions of $(\text{length})^{-1}$ and spatial volume has dimensions (length)^3 (see the next paragraph for the distinction between *physical* and *mathematical units*). The main advantage of having a dimensionless reduced Hamiltonian is that only such a reduced Hamiltonian can have a topological significance, and indeed, we will show that this renormalized reduced Hamiltonian is related to a topological invariant of M, namely, the σ-constant of M. Indeed, one of our main results (see Sect. 3) is that the σ-constant of M is related to the dimensionless reduced Hamiltonian by the following equation

$$\sigma(M) = -\tfrac{2}{3}\left(\inf_{(\gamma,p^{TT})\in P_{\text{reduced}}} H_{\text{reduced}}(\tau,\gamma,p^{TT})\right)^{2/3} . \tag{4}$$

On the other hand, we note that the value of a dimensionless Hamiltonian cannot be interpreted as an "energy".

Throughout this paper we use *mathematical units* wherein the coordinates, the metric coefficients, all curvatures, and volumes are dimensionless quantities. Thus, by our assertion that our new reduced Hamiltonian is dimensionless, we mean that it is dimensionless even when the metric variables are expressed in *physical units*, i.e., when the metric coefficients g_{ij} are taken with the dimensions of (length)^2 and the coordinates (x^i) are taken as being dimensionless. With these conventions, in *physical units*, the sectional curvature $K(g)$ and the scalar curvature $R(g)$ have dimensions of (length)$^{-2}$, the Ricci tensor $\text{Ric}(g)$ is dimensionless, and the volume $\text{vol}(M,g)$ has dimensions of (length)^3. Thus the physical metric g_{physical} is related to the mathematical metric $g_{\text{mathematical}}$ by

$$g_{\text{physical}} = \ell^2\, g_{\text{mathematical}} \tag{5}$$

where ℓ is a fixed positive constant with the dimension of (length); see also the remarks at the end of Sect. 2.4.

Regarding the range of the time coordinate t, we note that for manifolds of Yamabe type -1 that we consider here, the expected maximal ranges of the constant mean curvature τ is the interval $(-\infty,0)$ for which $\tau\to-\infty$ corresponds to a "crushing singular" big bang of vanishing spatial volume and $\tau\to 0$ corresponds to the limit of infinite volume expansion. The old time function $t=\frac{4}{3}\tau$ then corresponds to the same limits. The new time function $t=\frac{2}{3}\frac{1}{\tau^2}$, however, ranges in the interval $(0,\infty)$, vanishes at the big bang, and tends to positive infinity in the limit of infinite expansion.

We remark that to prove that a solution determined by Cauchy data prescribed at some initial time $t_0\in(0,\infty)$ actually exhausts the range $(0,\infty)$ and has the asymptotic volume properties suggested above is a difficult global existence problem that we shall not deal with here except to mention below some partial results that have been obtained elsewhere. Nevertheless, one of our main motivations for this work is the hope that reduction will lead to advances in the study of the global existence question for Einstein's equations.

We also remark that the model universes under study here could not cease expanding and begin to collapse since the onset of such a collapse would necessitate a "maximal" hypersurface having $\tau = 0$. But the Hamiltonian constraint $\mathcal{H}(g,\pi) = 0$ (see (7a) and (9a) below) would then yield the inequality $R(g) = (\pi \cdot \pi)\mu_g^{-2} \geq 0$ for the scalar curvature of the Riemannian spatial metric g and this is impossible to satisfy on a manifold of Yamabe type -1 (see Sect. 2.2).

Our results regarding the reduced Hamiltonian are described in more detail in Sect. 3. Here we briefly consider some additional background information and notation regarding the process of reduction. Our starting point for reduction is the Arnowitt-Deser-Misner *(ADM)* action for Einstein's equations [6], which takes the form

$$I_{ADM}(g,\pi) = \int_I \int_M \Big(\pi \cdot \partial_t g - N\mathcal{H}(g,\pi) - X \cdot \mathcal{J}(g,\pi)\Big)\, dt \tag{6}$$

where $I = [t_0, t_1] \subset \boldsymbol{R}$ is a closed interval, $\pi = ((\mathrm{tr}_g k)g - k)^\sharp \mu_g$ is the gravitational momentum, k is the second fundamental form, $tr_g k$ is the g-trace of k, μ_g is the volume element determined by g and the orientation of M, N is the lapse function and X is the shift vector field, and where the Hamiltonian density and momentum 1-form density are given by

$$\mathcal{H}(g,\pi) = \left(\pi \cdot \pi - \tfrac{1}{2}(\mathrm{tr}_g \pi)^2\right)\mu_g^{-1} - R(g)\mu_g \tag{7a}$$

$$\mathcal{J}(g,\pi) = 2(\delta_g \pi)^\flat \quad . \tag{7b}$$

In local coordinates, $\mathrm{tr}_g k = g^{ij}k_{ij}$, $\pi^{ij} = ((\mathrm{tr}_g k)g^{ij} - k^{ij})(\det g_{kl})^{1/2}$, $\mathrm{tr}_g \pi = g_{ij}\pi^{ij}$, $(\mu_g)_{123} = (\det g_{kl})^{1/2}$, $(\mu_g^{-1})_{123} = (\det g_{ij})^{-1/2}$, $R(g)$ is the scalar curvature of g, and center dot "$\cdot$" denotes double metric contraction, so that $\pi \cdot \partial g_t = \pi^{ij}\frac{\partial g_{ij}}{\partial t}$ and $\pi \cdot \pi = g_{ij}g_{kl}\pi^{ik}\pi^{jl}$. For the momentum 1$-$form density $\mathcal{J}(g,\pi)$, $\mathcal{J}_i(g,\pi) = 2((\delta_g\pi)^\flat)_i = -2g_{ij}\pi^{jk}{}_{|k}$, where vertical bar denotes covariant differentiation with respect to g, and where here center dot "$\cdot$" refers to the natural (i.e., non-metric contraction) of a vector and a 1-form density, $X \cdot \mathcal{J}(g,\pi) = X^i \mathcal{J}_i(g,\pi)$.

The main idea of reduction is that the *ADM* formulation of Einstein's equations is in "already-parameterized form", with a "super-Hamiltonian"

$$\int_M \Big(N\mathcal{H}(g,\pi) + X \cdot \mathcal{J}(g,\pi)\Big) \tag{8}$$

with lapse function N and shift vector field X that act as Lagrange multipliers. Consequently, the gravitational phase variables (g,π) must solve the Hamiltonian and divergence constraint equations

$$\mathcal{H}(g,\pi) = 0 \tag{9a}$$

$$\mathcal{J}(g,\pi) = 0 \quad . \tag{9b}$$

Thus one strives to find the true non-vanishing Hamiltonian of the theory by eliminating the constraints and imposing coordinate conditions (see [24],

p. 185, and [6], p. 231, for a discussion of the parametric form of the canonical equations).

Now suppose that M is of Yamabe type -1 and let $(\tau, \gamma, p^{TT}) \in \boldsymbol{R}^- \times P_{\text{reduced}}$ be reduced contact variables. Using the Choquet-Bruhat-Lichnerowicz-York [9] conformal method of solving the constraint equations, there exists a unique pair of *ADM* gravitational phase space variables (g, π) which have constant mean curvature and are related to the reduced variables (τ, γ, p^{TT}) by

$$(g, \pi) = \left(\varphi^4 \gamma,\ \varphi^{-4} p^{TT} + \tfrac{2}{3} \tau \varphi^2 \gamma^{-1} \mu_\gamma\right) \tag{10}$$

where the conformal factor $\varphi = \varphi(\tau, \gamma, p^{TT})$ is the unique positive solution of the Lichnerowicz equation, which we here write for an arbitrary γ (not necessarily for $\gamma \in \mathcal{M}_{-1}$)

$$\Delta_\gamma \varphi + \tfrac{1}{8} R(\gamma) \varphi + \tfrac{1}{12} \tau^2 \varphi^5 - \tfrac{1}{8} \left(p^{TT} \cdot p^{TT}\right) \mu_\gamma^{-2} \varphi^{-7} = 0 \ . \tag{11}$$

Here Δ_γ is the positive Laplacian $-\nabla_i \nabla^i$ with respect to γ, $p^{TT} \cdot p^{TT} = \gamma_{ik} \gamma_{jl} p^{TT^{ij}} p^{TT^{kl}}$ and $R(\gamma)$ is the scalar curvature of γ. Since the Choquet-Bruhat-Lichnerowicz-York method is conformally invariant and since M is of Yamabe type -1, each γ is pointwise conformally equivalent to a metric in $\mathcal{M}_{-1}$ and so we can now normalize without loss of generality by setting $R(\gamma) = -1$, thus assuring that the data (γ, p^{TT}) actually lies in P_{reduced}.

The resulting gravitational phase variables (g, π) are then solutions to the Hamiltonian and divergence constraint equations (9a) and (9b) and also satisfy the *CMC*-condition

$$\operatorname{tr}_g \pi = 2\tau\, \mu_g \ . \tag{12}$$

In terms of the reduced contact variables $(\tau, \gamma, p^{TT}) \in \boldsymbol{R}^- \times P_{\text{reduced}}$, the kinetic term in the *ADM* action (6) reduces to

$$\pi \cdot \partial_t g = p^{TT} \cdot \partial_t \gamma + \tfrac{4}{3} \partial_t \left(\tau \mu_{\varphi^4 \gamma}\right) - \tfrac{4}{3} \left(\frac{d\tau}{dt}\right) \mu_{\varphi^4 \gamma} \ .$$

Substituting this expression into the *ADM* action, using the fact that the constraints are now satisfied identically and so drop out of the action, and dropping a total time derivative, yields the reduced action in terms of the reduced variables (τ, γ, p^{TT}),

$$\begin{aligned} I_{\text{reduced}}(\tau, \gamma, p^{TT}) &= \int_I \int_M \pi \cdot \partial_t g\ dt \\ &= \int_I \int_M (p^{TT} \cdot \partial_t \gamma)\ dt - \tfrac{4}{3} \int_I \int_M \mu_{\varphi^4 \gamma} \frac{d\tau}{dt}\ dt \ . \end{aligned} \tag{13}$$

If in this expression we take $t = \frac{4}{3}\tau$ as a temporal coordinate condition, then the reduced Hamiltonian density is given in terms of the canonical contact variables (τ, γ, p^{TT}), by

$$\mathcal{H}^{\text{old}}_{\text{reduced}}(\tau, \gamma, p^{TT}) = \varphi^6 \mu_\gamma = \mu_{\varphi^4 \gamma} = \mu_g$$

and the reduced Hamiltonian is given by

$$\begin{aligned} H^{\text{old}}_{\text{reduced}}(\tau,\gamma,p^{TT}) &= \int_M \mathcal{H}^{\text{old}}_{\text{reduced}}(\tau,\gamma,p^{TT}) = \int_M \varphi^6\mu_\gamma \\ &= \int_M \mu_{\varphi^4\gamma} = \int_M \mu_g = \text{vol}(M,g)\ . \end{aligned} \tag{14}$$

The dependence of $\mu_g = \varphi^6\mu_\gamma$ on the variables (τ,γ,p^{TT}) is through the function $\varphi = \varphi(\tau,\gamma,p^{TT})$, which, being the solution of the elliptic Lichnerowicz equation (11) is a non-local function of its arguments (τ,γ,p^{TT}).

In this paper we re-normalize our previous work in order to introduce a dimensionless reduced Hamiltonian. Thus, instead of taking the time function t to be proportional to the mean curvature τ, we let $t = \frac{2}{3\tau^2}$ so that $\frac{4}{3}\frac{d\tau}{dt} = -\tau^3$. The reduced Hamiltonian density is then given by

$$\mathcal{H}^{\text{new}}_{\text{reduced}}(\tau,\gamma,p^{TT}) = \mathcal{H}_{\text{reduced}}(\tau,\gamma,p^{TT}) = -\tau^3\varphi^6\mu_\gamma = -\tau^3\mu_{\varphi^4\gamma} = -\tau^3\mu_g$$

and the reduced Hamiltonian is given by

$$\begin{aligned} H^{\text{new}}_{\text{reduced}}(\tau,\gamma,p^{TT}) &= H_{\text{reduced}}(\tau,\gamma,p^{TT}) = -\tau^3 \int\limits_M \varphi^6\mu_\gamma \\ &= -\tau^3 \int\limits_M \mu_{\varphi^4\gamma} = -\tau^3 \int\limits_M \mu_g = -\tau^3\text{vol}(M,g) \end{aligned} \tag{15}$$

where again the conformal factor $\varphi = \varphi(\tau,\gamma,p^{TT})$ is the unique positive solution of the Lichnerowicz equation (11), and where from now on we suppress the "new" superscript.

We remark that the fact that our Hamiltonian reduction process results in a formulation of dynamics that involves a contact manifold, as opposed to a simpler symplectic one, is forced upon us by the essential dependence of the conformal factor $\varphi = \varphi(\tau,\gamma,p^{TT})$ upon τ and seems to be unavoidable. That the Hamiltonian depends essentially upon time, and not merely through the overall factor of $-\tau^3$, results from the inevitable volume expansion (or contraction in the time-reversed case) of our model universes. In fact, we shall see in Sects. 3 and 4 that the decaying factor $-\tau^3$ precisely cancels the increase in volume only for very special, and in fact flat, spacetimes which arise if and only if M is hyperbolic.

In this paper we shall study some of the remarkable properties of H_{reduced} relating to its dependence on time, its critical points, and its relation to the σ-constant of the underlying manifold M.

2 Some Mathematical Background

In this section we briefly review some of the mathematical background material that we will use regarding the reduced phase space and also regarding

3-manifold topology. Our goal is to describe in a fairly specific manner the 3-manifold topologies for which our analysis applies. For details of the results summarized here, see [13] and [14].

2.1 The Teichmüller Space of Conformal Structures of a Manifold

Let M be a compact connected smooth (C^∞) oriented n-manifold without boundary, $n \geq 2$. The main spaces that we shall consider are the following:

$\mathcal{M}$ $=$ Riem(M) $=$ the space of smooth (C^∞) Riemannian metrics on M

$\mathcal{D}$ $=$ Diff(M) $=$ the group of smooth diffeomorphisms of M

$\mathcal{D}^+$ $=$ Diff$^+(M)$ $=$ the subgroup of orientation-preserving diffeomorphisms of M

$\mathcal{D}_0$ $=$ Diff$_0(M)$ $=$ the connected component of the identity diffeomorphism of M

$\mathcal{P}$ $=$ Pos(M) $=$ the space of smooth real-valued positive functions on M

Occasionally, when more than one manifold is being discussed, we will include the underlying manifold in our notation, writing, e.g., $\mathcal{M} = \mathcal{M}(M)$, or $\mathcal{D} = \mathcal{D}(M)$, for clarity.

For a Riemannian metric $g \in \mathcal{M}$, we let

$K(g)$ $=$ the sectional curvature of g

Ric(g) $=$ the Ricci curvature tensor of g

$R(g)$ $=$ the scalar curvature tensor of g

μ_g $=$ the unique volume element on M determined by g and the orientation of M

$$\mathrm{vol}(M, g) = \int_M \mu_g = \text{ the volume of } (M, g) \ .$$

Our sign conventions on the curvature tensors are as in [23].

The group $\mathcal{D}$ acts on $\mathcal{M}$ by pullback and the group $\mathcal{P}$ acts on $\mathcal{M}$ by pointwise multiplication. The resulting orbit space

$$\mathcal{S} = \frac{\mathcal{M}}{\mathcal{D}}$$

of Riemannian geometries is defined as *superspace*; the resulting orbit space

$$\frac{\mathcal{M}}{\mathcal{P}}$$

of pointwise conformal structures is defined as *pointwise conformal superspace*; and the resulting orbit space

$$\mathcal{C} = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}}$$

of conformal geometries is defined as *conformal superspace*. We are also interested in the $\mathcal{D}^+$-restricted counterparts of superspace and conformal superspace; namely, $\mathcal{D}^+$-*restricted superspace*

$$\mathcal{S}^+ = \frac{\mathcal{M}}{\mathcal{D}^+}$$

and $\mathcal{D}^+$-*restricted conformal superspace*

$$\mathcal{C}^+ = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}^+} \ ,$$

as well as in the $\mathcal{D}_0$-restricted counterparts of superspace and conformal superspace; namely, $\mathcal{D}_0$-*restricted superspace*

$$\mathcal{S}_0 = \frac{\mathcal{M}}{\mathcal{D}_0}$$

and $\mathcal{D}_0$-*restricted conformal superspace*

$$\mathcal{C}_0 = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}_0} \ .$$

When $n = 2$, let $M = \Sigma_p$ denote a compact connected oriented surface of genus $p \geq 0$, so that $\Sigma_0 = \boldsymbol{S}^2 =$ the 2-sphere, $\Sigma_1 = \boldsymbol{T}^2 =$ the 2-torus, and Σ_p, $p \geq 2$, is a “higher genus surface”. When $n = 2$, $\mathcal{D}^+$-restricted conformal superspace $\mathcal{C}^+(\Sigma_p)$ is the *Riemann moduli space* $\mathcal{R}_p$ and $\mathcal{D}_0$-restricted conformal superspace $\mathcal{C}_0(\Sigma_p)$ is the *Teichmüller space* $\mathcal{T}_p$ of Σ_p (see [16]).

Thus, for $n = 2$, $p \geq 0$, we have

$$\mathcal{T}_p = \mathcal{T}(\Sigma_p) = \mathcal{C}_0(\Sigma_p) \quad \text{and} \quad \mathcal{R}_p = \mathcal{R}(\Sigma_p) = \mathcal{C}^+(\Sigma_p) \ .$$

Analogously, for $n \geq 3$, we shall refer to $\mathcal{C}^+(M)$ as the *Riemann moduli space of conformal structures on* M and $\mathcal{C}_0(M)$ as the *Teichmüller space of conformal structures on* M, and write

$$\mathcal{T}_M = \mathcal{T}(M) = \mathcal{C}_0(M) \quad \text{and} \quad \mathcal{R}_M = \mathcal{R}(M) = \mathcal{C}^+(M) \ .$$

To put this terminology into further focus, consider the space

$$\mathcal{M}_{-1} = \{g \in \mathcal{M} \mid R(g) = -1\} \subset \mathcal{M}$$

of Riemannian metrics with constant scalar curvature -1. In [12] it is shown that $\mathcal{M}_{-1}$ is a smooth submanifold of $\mathcal{M}$.

If $n = 2$ and genus $p \geq 2$, then the space of pointwise conformal structures on Σ_p is smoothly diffeomorphic to $\mathcal{M}_{-1}$,

$$\frac{\mathcal{M}(\Sigma_p)}{\mathcal{P}(\Sigma_p)} \approx \mathcal{M}_{-1}(\Sigma_p) \ . \tag{16}$$

Thus, the Riemann moduli space of Σ_p can be represented by

$$\mathcal{R}_p = \mathcal{C}^+(\Sigma_p) = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}^+} \approx \frac{\mathcal{M}_{-1}}{\mathcal{D}^+} \tag{17}$$

and the Teichmüller space of Σ_p can be represented by

$$\mathcal{T}_p = \mathcal{C}_0(\Sigma_p) = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}_0} \approx \frac{\mathcal{M}_{-1}}{\mathcal{D}_0} \ . \tag{18}$$

For $n \geq 3$, we shall define the Yamabe type of a manifold below. In this paper we shall require that M be of Yamabe type -1 and we shall see that this requirement is analogous in the case that $n = 2$ to the requirement that $p \geq 2$. Indeed, if $n \geq 3$ and M is of Yamabe type -1, then the representations in (16), (17), and (18) all remain valid, so that

$$\frac{\mathcal{M}}{\mathcal{P}} \approx \mathcal{M}_{-1} \ , \tag{19}$$

and thus the Riemann moduli space of M can be represented by

$$\mathcal{R}_M = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}^+} \approx \frac{\mathcal{M}_{-1}}{\mathcal{D}^+} \tag{20}$$

and the Teichmüller space of M can be represented by

$$\mathcal{T}_M = \frac{\mathcal{M}/\mathcal{P}}{\mathcal{D}_0} \approx \frac{\mathcal{M}_{-1}}{\mathcal{D}_0} \ . \tag{21}$$

2.2 The Yamabe Type of a Manifold

Recall that a *space form* is a complete connected Riemannian manifold (M, g) of constant sectional curvature. The space form is *spherical* if the sectional curvature of g is positive, *Euclidean* if the sectional curvature is zero, and *hyperbolic* if the sectional curvature is negative.

A connected differentiable manifold M (with or without a Riemannian structure) is *spherical* (respectively, *Euclidean*; respectively, *hyperbolic*) if M is diffeomorphic to the underlying manifold of a spherical space form (respectively, a Euclidean space form; respectively, a hyperbolic space form).

Thus, for example, a connected manifold M is *hyperbolic* if there exists a hyperbolic Riemannian metric on M, whereas M is *non-hyperbolic* if no Riemannian metric on M is hyperbolic.

Now let M be a compact connected n-manifold, $n \geq 3$. We introduce the following terminology.

1. M is of *Yamabe-type* -1, written $Y(M) = -1$, if M admits no metric with $R(g) = 0$;
2. M is of *Yamabe-type* 0, written $Y(M) = 0$, if M admits a metric with $R(g) = 0$, but no metric with $R(g) = 1$; and

3. M is of *Yamabe-type* $+1$ (*or* 1), written $Y(M) = 1$, if M admits a metric with $R(g) = 1$.

The definition of Yamabe type partitions the class of compact connected n-manifolds, $n \geq 3$, into three classes that are mutually exclusive and exhaustive (see [13] for details).

We note in particular that spherical manifolds are a subset of manifolds of Yamabe type 1, compact Euclidean manifolds are a subset of manifolds of Yamabe type 0 since such manifolds cannot have a metric with constant positive scalar curvature, and compact hyperbolic manifolds are a subset of manifolds of Yamabe type -1 since such manifolds cannot have a metric with zero scalar curvature (see [25], p. 306). However, we remark that compactness is crucial in this characterization inasmuch as the manifold $\boldsymbol{R}^n$ supports both a flat and a hyperbolic metric and thus is both a Euclidean and a hyperbolic manifold.

2.3 Some 3-Manifold Topology

We now restrict our attention to 3-manifolds. Our goal is to relate the topology of 3-manifolds to their Yamabe type. To simplify the discussion, we assume that the following standard conjecture of 3-manifold topology is true:

Elliptization Conjecture: Every compact connected orientable 3*-manifold* M *with finite fundamental group is a spherical manifold (i.e., is diffeomorphic to a spherical space form).*

This conjecture is part of Thurston's [41] geometrization program. The elliptization conjecture is equivalent to the Poincaré conjecture and a conjecture asserting that the only free actions of finite groups on $\boldsymbol{S}^3$ are the standard orthogonal actions (see also [40] for further information regarding Thurston's program).

Recall that a $K(\boldsymbol{\pi}, 1)$-manifold is a manifold M whose first homotopy group $\pi_1(M) = \boldsymbol{\pi}$ and all of whose higher homotopy groups vanish (equivalently, the universal covering space of M is contractible). Such a manifold is also said to be *aspherical*.

Let M be a non-trivial ($\not\approx \boldsymbol{S}^3$) compact connected orientable 3-manifold. Assume that the elliptization conjecture is true. Then the Kneser-Milnor ([22], [28]) prime decomposition theorem asserts that M is diffeomorphic to a finite connected sum of the following form,

$$\begin{aligned} M \approx & \underbrace{\boldsymbol{S}^3/\boldsymbol{\Gamma}_1 \# \ldots \# \boldsymbol{S}^3/\boldsymbol{\Gamma}_k}_{\text{spherical factors}} \# \underbrace{(\boldsymbol{S}^1 \times \boldsymbol{S}^2)_1 \# \ldots \# (\boldsymbol{S}^1 \times \boldsymbol{S}^2)_l}_{\text{handles (or wormholes)}} \\ & \# \underbrace{K(\boldsymbol{\pi}_1, 1) \# \ldots \# K(\boldsymbol{\pi}_m, 1)}_{\text{aspherical factors}} \end{aligned} \tag{22}$$

where

1. k, l, and m are non-negative integers (if any of the integers k, l, or m are zero, then terms of that type do not appear);
2. if $k \geq 1$, then for each i, $1 \leq i \leq k$, $\boldsymbol{\Gamma}_i \subset \boldsymbol{SO}(4)$ is a finite non-trivial subgroup of $\boldsymbol{SO}(4)$ acting freely and orthogonally on $\boldsymbol{S}^3$, so that each resulting factor $\boldsymbol{S}^3/\boldsymbol{\Gamma}_i$ is a spherical space form;
3. if $m \geq 1$, then for each j, $1 \leq j \leq m$, each aspherical factor is a $K(\boldsymbol{\pi}_j, 1)$-manifold with infinite fundamental group $\pi_1(K(\boldsymbol{\pi}_j, 1)) = \boldsymbol{\pi}_j$;

and where the summands in (22) are uniquely determined up to order and diffeomorphism.

Assuming the elliptization conjecture, using the prime decomposition theorem, and using results of Gromov-Lawson ([17], [18], [19]) and Schoen-Yau ([36], [37], [38]), we can give a characterization of the structure of the manifolds of the three Yamabe types (see [13] for details).

Let M be a compact connected orientable 3-manifold. Then,

1. $Y(M) = -1$ if and only if either M is a non-Euclidean $K(\pi, 1)$-manifold (i.e., M is a "stand-alone" $K(\pi, 1)$-manifold that does not admit a flat Riemannian metric), or M is a composite manifold such that in its connected sum decomposition (22) there is at least one $K(\pi, 1)$-factor (which may be Euclidean), i.e., $M \approx M' \# K(\boldsymbol{\pi}, 1)$, where $M' \not\approx \boldsymbol{S}^3$;
2. $Y(M) = 0$ if and only M is Euclidean, and thus M is diffeomorphic to one of six orientable Euclidean space forms (see [42] for a description of these space forms);
3. $Y(M) = 1$ if and only M is diffeomorphic to a finite connected sum of spherical space forms and handles, i.e.,

$$M \approx \boldsymbol{S}^3/\boldsymbol{\Gamma}_1 \# \ldots \# \boldsymbol{S}^3/\boldsymbol{\Gamma}_k \# (\boldsymbol{S}^1 \times \boldsymbol{S}^2)_1 \# \ldots \# (\boldsymbol{S}^1 \times \boldsymbol{S}^2)_l \tag{23}$$

Our main interest in this paper is in compact connected orientable 3-manifolds with $Y(M) = -1$. Thus, assuming that the elliptization conjecture is true, the structure of these 3-manifolds is given by 1. above.

2.4 Conformal Geometry and the σ-Constant

We now discuss some results from conformal geometry that we shall need, specialized to 3-manifolds of Yamabe type -1. We then discuss the σ-constant of M, $\sigma(M)$, a fundamental topological invariant for M. Further information regarding this material is given in [2], [26], and [39].

Let M be a compact connected oriented n-manifold, $n \geq 3$. The *Yamabe functional* is defined by

$$Y : \mathcal{M} \longrightarrow \boldsymbol{R}\,, \quad g \longmapsto \frac{\int_M R(g)\mu_g}{(\int_M \mu_g)^{(n-2)/n}} \ . \tag{24}$$

If $\psi \in \mathcal{P} = \mathrm{Pos(M)}$, then

$$Y\left(\psi^{4/(n-2)}g\right) = \frac{\int_M (4(\frac{n-1}{n-2})\|d\psi\|_g^2 + R(g)\psi^2)\mu_g}{(\int_M \psi^{2n/(n-2)}\mu_g)^{(n-2)/n}} \ . \tag{25}$$

Specializing to the case $n = 3$,

$$Y(g) = \frac{\int_M R(g)\mu_g}{(\int_M \mu_g)^{1/3}} \ , \tag{26}$$

and if $\psi \in \mathcal{P} = \mathrm{Pos(M)}$, then

$$Y(\psi^4 g) = \frac{\int_M (8\|d\psi\|_g^2 + R(g)\psi^2)\mu_g}{(\int_M \psi^6 \mu_g)^{1/3}} \ . \tag{27}$$

Fixing g and letting ψ vary, by Hölder's inequality this functional as a function of ψ is bounded below. One defines the *Yamabe invariant $y(g)$ of g* (or the *Yamabe constant*), by

$$y(g) = \inf_{g' \in \langle g \rangle} Y(g') = \inf_{\psi \in \mathcal{P}} Y(\psi^4 g) = \inf_{\psi \in \mathcal{P}} \frac{\int_M (8\|d\psi\|_g^2 + R(g)\psi^2)\mu_g}{(\int_M \psi^6 \mu_g)^{1/3}} \tag{28}$$

where $\langle g \rangle = \{\psi^4 g \mid \psi \in \mathcal{P}\}$ is the pointwise conformal class of g.

The Yamabe invariant is a conformal invariant of g, depending only on the conformal class of g. Thus, if $\psi \in \mathcal{P}$, then $y(\psi^4 g) = y(g)$, so that y passes to the orbit space of pointwise conformal structures (and indeed to conformal structures)

$$w : \mathcal{M}/\mathcal{P} \longrightarrow \boldsymbol{R}\,; \quad \langle g \rangle \longmapsto w(\langle g \rangle) = y(g')\,, \quad g' \in \langle g \rangle \ ,$$

which is *the Yamabe invariant of the conformal class* $\langle g \rangle$.

If $(\boldsymbol{S}^3, g_1)$ denotes the standard 3-sphere with unit radius, then the Yamabe invariant satisfies the following bound ([7]),

$$y(g) = y(M, g) \leq y(\boldsymbol{S}^3, g_1) \ . \tag{29}$$

The Yamabe invariant $y(g)$ is an infimum within a fixed conformal class. Using the bound given by (29), it is natural to consider the supremum of the Yamabe invariants over all conformal classes. Thus one defines the σ-constant of M, $\sigma(M)$, by taking the supremum of the Yamabe invariant over all conformal classes of M,

$$\sigma(M) = \sup_{\langle g\rangle \in \mathcal{M}/\mathcal{P}} w\Big(\langle g\rangle\Big) = \sup_{\langle g\rangle \in \mathcal{M}/\mathcal{P}} \Big(\inf_{g' \in \langle g\rangle} Y(g')\Big) \tag{30}$$

where the supremum exists by (29). Thus the σ-constant is defined by a minimax process (first infimum, then supremum) analogous to the minimax process (first supremum, then infimum) used in Morse theory (see [31], [32], and [33]).

Since $y(g) = w(\langle g\rangle)$ is conformally invariant, (30) reduces to

$$\sigma(M) = \sup_{\langle g\rangle \in \mathcal{M}/\mathcal{P}} w\Big(\langle g\rangle\Big) = \sup_{g \in \mathcal{M}} y(g) = \sup_{g \in \mathcal{M}} \Big(\inf_{\psi \in \mathcal{P}} Y(\psi^4 g)\Big) \ . \tag{31}$$

The main idea behind the introduction and use of the σ-constant is to use it as a tool to search for Einstein metrics on general compact manifolds ([2]). The usual procedure is to seek critical points of the total scalar curvature functional restricted to metrics with unit volume,

$$R^1_T : \mathcal{M}^1 \longrightarrow \boldsymbol{R} \ , \qquad g \longmapsto \int_M R(g)\mu_g \ , \tag{32}$$

where

$$\mathcal{M}^1 = \{\, g \in \mathcal{M} \mid \mathrm{vol}(M, g) = \int_M \mu_g = 1 \,\} \ , \tag{33}$$

for such critical points are necessarily Einstein metrics.

This approach has two difficulties. Firstly, the functional R^1_T is bounded neither above nor below. Secondly, any critical point of R^1_T has both infinite index and co-index. Thus, the usual methods of obtaining the existence of critical points cannot be applied.

Thus one introduces a minimax procedure in order to obtain the critical values of R^1_T, and then, hopefully, the critical points. The minimax procedure to obtain $\sigma(M)$ has two steps. First minimize the Yamabe functional in a fixed conformal class and then maximize over all conformal classes. The first step of this procedure, corresponding to the *min* part, is the Yamabe problem and has been solved (see [34] and [39] and the references therein). Indeed, if the Yamabe invariant $y(g)$ is achieved by g',

$$Y(g') = y(g) = \inf_{\psi \in \mathcal{P}} Y(\psi^4 g) \ , \tag{34}$$

then g' must necessarily have constant scalar curvature.

The second step of the minimax procedure, maximizing over the conformal classes is considerably more difficult and has not been solved. However, it is known that if the σ-constant of M is ≤ 0 and is achieved by g',

$$Y(g') = y(g') = \sigma(M) \ , \tag{35}$$

then g' is necessarily an Einstein metric and *a fortiori* has constant scalar curvature. In the case that $\sigma(M) > 0$, it is conjectured that if g' realizes $\sigma(M)$, then g' is an Einstein metric, but this remains unknown.

In this regard we remark that a constant curvature metric (unique up to isometry and homothety) on $\boldsymbol{S}^n$ achieves the σ-constant of $\boldsymbol{S}^n$, and that any flat metric on $\boldsymbol{T}^n$ achieves the σ-constant of $\boldsymbol{T}^n$. However, it is still unknown, but conjectured to be true, that the σ-constant of a hyperbolic manifold M is achieved by a hyperbolic metric on M (which by Mostow rigidity is unique up to isometry and homothety). We shall see that this conjecture is equivalent to a conjecture of ours that the local minimum of our reduced Hamiltonian is a global minimum (see Sects. 5 and 6).

We also remark that (24) is defined for $n = 2$. In this case, by the Gauss-Bonnet theorem, $Y(g) = 4\pi\chi(\Sigma_p)$ and so

$$\sigma(\Sigma_p) = y(g) = 4\pi\chi(\Sigma_p) = 8\pi(1-p) \ . \tag{36}$$

Thus in some sense one can think of the σ-constant for $n \geq 3$ as a generalization of the Euler characteristic for $n = 2$; see also Sect. 6.

Now we specialize to the case where $Y(M) = -1$. Since the Yamabe invariant $y(g)$ depends only upon the conformal class of g, we may replace g by any metric conformal to g. Since $Y(M) = -1$, every metric g is pointwise conformally equivalent to a metric $\gamma \in \mathcal{M}_{-1}$. Thus for g and γ, $y(g) = y(\gamma)$. In this case the infimum of (28) is actually achieved by a constant $\psi = c > 0$ and then the Yamabe invariant is given by

$$y(g) = y(\gamma) = \frac{-\int_M c^2 \mu_\gamma}{(\int_M c^6 \mu_\gamma)^{1/3}} = -\left(\int_M \mu_\gamma\right)^{2/3} \ . \tag{37}$$

Thus, if $Y(M) = -1$, every conformal class is uniquely represented by a metric in $\mathcal{M}_{-1}$, so from (31) and (37), we have the following fundamental equation for $\sigma(M)$,

$$\begin{aligned} \sigma(M) &= \sup_{g\in\mathcal{M}} y(g) = \sup_{\gamma\in\mathcal{M}_{-1}} y(\gamma) \\ &= \sup_{\gamma\in\mathcal{M}_{-1}} \left(-\left(\int_M \mu_\gamma\right)^{2/3} \right) = -\inf_{\gamma\in\mathcal{M}_{-1}} \left(\int_M \mu_\gamma\right)^{2/3} \\ &= -\left(\inf_{\gamma\in\mathcal{M}_{-1}} \int_M \mu_\gamma\right)^{2/3} = -\left(\inf_{\gamma\in\mathcal{M}_{-1}} \mathrm{vol}(M,\gamma)\right)^{2/3} \ . \end{aligned} \tag{38}$$

Thus, if $Y(M) = -1$, the minimax process defining the σ-constant is replaced by a simpler *min* (i.e., infimum) process since the first step of the process, taking the infimum, is actually achieved by a metric $\gamma \in \mathcal{M}_{-1}$, and since the second step of the process, taking the supremum, is converted to an infimum since $y(\gamma) = -\left(\int_M \mu_\gamma\right)^{2/3} < 0$.

Thus, *if M is a compact connected oriented 3-manifold with $Y(M) = -1$, then $\sigma(M)$ is expressed in terms of the infimum of the volume of Riemannian metrics restricted to the submanifold $\mathcal{M}_{-1}$*. Moreover, in Sect. 4, we shall show that *this infimum can be attained only if γ is a hyperbolic metric.*

We shall return to these fundamental results in Sect. 5. Here we make two further remarks. Firstly, if M is a compact connected oriented n-manifold, $n \geq 3$, with $Y(M) = -1$, then, more generally,

$$\sigma(M) = -\left(\inf_{\gamma \in \mathcal{M}_{-1}} \mathrm{vol}(M, \gamma)\right)^{2/n} . \tag{39}$$

Secondly, regarding the units of the Yamabe functional $Y(g)$ (see the discussion of units in Sect. 1), $n \geq 3$, we remark that in physical units $Y(g)$ is dimensionless. Indeed, letting ℓ denote a fixed positive constant with dimension of (length), then

$$\begin{aligned}\text{Physical Units}\left(\frac{\int_M R(g)\mu_g}{(\int_M \mu_g)^{(n-2)/n}}\right) &= \frac{\ell^{-2} \cdot \mathrm{vol}}{\mathrm{vol}^{(n-2)/n}} \\ &= \ell^{-2} \cdot \mathrm{vol}^{2/n} = \ell^{-2} \cdot (\ell^n)^{2/n} = \ell^0 = 1 \ .\end{aligned} \tag{40}$$

Thus $y(g)$, $w(\langle g \rangle)$, and $\sigma(M)$ are all *ipso facto* dimensionless.

We note, however, that the reason that (39) (and (38)) appear to have physical units ℓ^2 is that when we set $R(g) = -1$ in (37), the -1 that appears is only numerically -1 but in physical units has dimensions ℓ^{-2}. Similarly, in physical units the $- = -1$ that appears in (38) and (39) is numerically -1 but has dimensions ℓ^{-2} and thus in physical units $\sigma(M)$ is dimensionless.

3 Properties of the Reduced Hamiltonian

Throughout this section we assume that M is a smooth (C^∞) compact connected oriented 3-manifold without boundary *such that M is of Yamabe type* -1.

As discussed in Sect. 2.2, a subclass of manifolds of Yamabe type -1 are the hyperbolic manifolds. If M is hyperbolic, then, by Mostow rigidity, up to isometry, there is a unique hyperbolic metric $\tilde{\gamma}$ on M with sectional curvature $K(\tilde{\gamma}) = -\frac{1}{6}$ and scalar curvature $R(\tilde{\gamma}) = -1$. Then, correspondingly, up to isometry, there is a unique $(\tilde{\gamma}, 0) \in P_{\text{reduced}}$.

Our aim in this paper is to establish the following properties of the dimensionless reduced Hamiltonian $H_{\text{reduced}}(\tau, \gamma, p^{TT}) = -\tau^3 \int_M \varphi^6 \mu_\gamma$ discussed in Sect. 1.

1. If M is a hyperbolic manifold, then $H_{\text{reduced}}(\tau, \gamma, p^{TT})$ is a monotonically decreasing function of t unless $p^{TT} = 0$ and $\gamma = \tilde{\gamma}$ is hyperbolic (with $K(\tilde{\gamma}) = -\frac{1}{6}$). At $(\gamma, p^{TT}) = (\tilde{\gamma}, 0)$, $H_{\text{reduced}}(\tau, \tilde{\gamma}, 0)$ is constant in time.

2. If M is a non-hyperbolic manifold (but of Yamabe type -1), then $H_{\text{reduced}}(\tau,\gamma,p^{TT})$ is globally monotonically decreasing in t.
3. If M is a hyperbolic manifold and $\tau \in (-\infty,0)$ is fixed, then $H_{\text{reduced}}(\tau,\gamma,p^{TT})$ has a critical point at $(\tilde{\gamma},0)$, which is unique up to isometry. Moreover, $(\tilde{\gamma},0)$ is a strict local minima of $H_{\text{reduced}}(\tau,\gamma,p^{TT})$ in the non-isometric directions.
4. If M is a non-hyperbolic manifold (but of Yamabe type -1) and $\tau \in (-\infty,0)$ is fixed, then $H_{\text{reduced}}(\tau,\gamma,p^{TT})$ has no critical points.
5. If M is of Yamabe type -1 and $\tau \in (-\infty,0)$ is fixed, then the σ-constant of M is related to the reduced Hamiltonian as follows:

$$\sigma(M) = -\tfrac{2}{3}\left(\inf_{(\gamma,p^{TT})\in P_{\text{reduced}}} H_{\text{reduced}}(\tau,\gamma,p^{TT})\right)^{2/3} \tag{41}$$

or, alternately,

$$\inf_{(\gamma,p^{TT})\in P_{\text{reduced}}} H_{\text{reduced}}(\tau,\gamma,p^{TT}) = \left(-\tfrac{3}{2}\sigma(M)\right)^{3/2} . \tag{42}$$

Remarks:

1. If M is a hyperbolic manifold, then corresponding to the unique hyperbolic critical point $(\tilde{\gamma},0)$ are the constant mean curvature *ADM* gravitational variables (see (51))

$$(g,\pi) = \left(g,\ \tfrac{2}{3}\tau g^{-1}\mu_g\right) = \left(\tfrac{3}{2}\tau^{-2}\tilde{\gamma},\ -\left(\tfrac{2}{3}\right)^{1/2}\tilde{\gamma}^{-1}\mu_{\tilde{\gamma}}\right) .$$

The corresponding solutions of Einstein's equations are flat, admit a global timelike homothetic vector field, and are isometric to quotients of the interior of the future light cone of a point of Minkowski spacetime by co-compact discrete subgroups of the proper orthochronous Lorentz group which fixes that point (see (48)).
2. If M is a hyperbolic manifold, then $(\tilde{\gamma},0)$ is a strict local minimum of $H_{\text{reduced}}(\tau,p^{TT})$ and $H_{\text{reduced}}(\tau,\gamma,p^{TT})$ has no other critical points (up to isometry). We conjecture that $(\tilde{\gamma},0)$ is a strict global minimum which is equivalent to the conjecture that the σ-constant of M is realized by the hyperbolic geometry on M. If M is not hyperbolic, then the σ-constant is never realized by a metric on M but is only approached as a limit.

In other dimensions the most natural definition of the reduced Hamiltonian (for a *CMC* slicing) is $H_{\text{reduced}} = |\tau|^n \int_M \mu_g$. For any dimension $n \neq 1$ (which would be a degenerate case) this form of H_{reduced} follows from a choice of time function $t = \frac{\text{constant}}{\tau^{n-1}}$. Moreover, in other dimensions, all of these properties of H_{reduced} have analogues in $n+1$ gravity.

For example, in $2+1$ gravity where the role of the σ-constant of M is played by the Euler characteristic $\chi(\Sigma_p)$ of a surface Σ_p, $p \geq 2$, the conjecture that the critical points of H_{reduced} are in fact global minima is known to be true (see also Sect. 6). In fact, there is already a global existence theorem for vacuum $2+1$ gravity on manifolds of the type $\boldsymbol{R}^- \times \Sigma_p$ which shows that every solution, when extended to its maximal Cauchy development exhausts the interval $\tau \in (-\infty, 0)$ and has asymptotic properties for area(Σ_p, g) of the type suggested above for vol(M, g), i.e., the area collapses to zero at the big bang where $\tau \to -\infty$ and expands to infinity as $\tau \to 0$ (see [5]). In vacuum $2+1$ gravity all solutions are flat, but the special solutions with a global timelike homothety which arise from the critical points of H_{reduced} in that case are represented by the zero section of $T^*(\mathcal{M}_{-1}(\Sigma_p))/\mathcal{D}_0(\Sigma_p))$ which, in turn, is diffeomorphic to the Teichmüller space $\mathcal{T}_p = \mathcal{T}(\Sigma_p) = \mathcal{M}_{-1}(\Sigma_p)/\mathcal{D}_0(\Sigma_p)$ of Σ_p, where the points of Teichmüller space are represented by the $\mathcal{D}_0(\Sigma_p)$-isometry classes of hyperbolic metrics on Σ_p (see Sect. 2.1). However, in $3+1$ gravity, because of Mostow rigidity, when M is of hyperbolic type, then up to isometry there is a unique and hence isolated hyperbolic critical point.

In $n+1$ dimensions with $n \geq 3$, the condition that (γ, p^{TT}) be a critical point of H_{reduced} is that $p^{TT} = 0$ and that g be an Einstein metric (with negative Einstein constant). For $n = 3$, this leads to the requirement that γ be a hyperbolic metric, and hence by Mostow rigidity leads to the uniqueness of the critical point.

For $n \geq 4$, the criticality condition that γ be an Einstein metric is not as rigid as in dimension $n = 3$, and indeed does not lead to a unique or isolated critical point. When $n = 2$, although the critical condition does imply constant curvature, there is still a Teichmüller space of deformations possible. *Thus only when $n = 3$ do we get a unique isolated critical point of* H_{reduced} (see also the remarks preceding (49)).

4 Proofs of the Properties of H_{reduced}

To show that H_{reduced} is monotonically decreasing we compute

$$\begin{aligned} \frac{d}{dt} H_{\text{reduced}} &= \frac{d}{dt}\Big(-\tau^3 \int_M \mu_g\Big) \\ &= -3\tau^2 \frac{d\tau}{dt} \int_M \mu_g + \tau^3 \int_M \Big(\tfrac{1}{2} N (tr_g \pi) \mu_g^{-1} - \text{div}_g X\Big) \mu_g \\ &= \tfrac{9}{4}\tau^5 \int_M \mu_g + \tau^4 \int_M N \mu_g \end{aligned} \tag{43}$$

where N is the lapse function, X is the shift vector field, and where we have used the gauge condition $t = \frac{2}{3}\frac{1}{\tau^2}$ and the compactness of M to eliminate the term involving the divergence of X, $\text{div}_g X = X^i{}_{|i}$. To evaluate the integral of N we need the elliptic equation which determines N for the chosen *CMC* slicing

$$\frac{\partial \tau}{\partial t} = -\tfrac{3}{4}\tau^3 = \Delta_g N + \Big(\pi \cdot \pi - \tfrac{1}{4}(tr_g \pi)^2\Big) \mu_g^{-2} N$$

$$= \Delta_g N + \left((\pi^{TT} \cdot \pi^{TT})\mu_g^{-2} + \tfrac{1}{3}\tau^2\right)N \tag{44}$$

where $\pi^{TT} = \pi - \frac{1}{3}g^{-1}(\mathrm{tr}_g\pi) = \pi - \frac{2}{3}\tau g^{-1}\mu_g$. Integrating (44) over M and substituting the result into (43) yields

$$\frac{d}{dt}H_{\text{reduced}} = -3\tau^2 \int_M N\left(\pi^{TT} \cdot \pi^{TT}\right)\mu_g^{-1} \ . \tag{45}$$

We remark that this equation was shown to hold in arbitrary dimensions by Alan Rendall (unpublished). A standard maximum principle argument applied to (44) shows that $N > 0$ and thus, since $\tau \neq 0$, we obtain $\frac{d}{dt}H_{\text{reduced}} \leq 0$ with equality holding if and only if $\pi^{TT} = 0$. Thus, if $\pi^{TT} \neq 0$, then H_{reduced} is monotonically decreasing. We also remark that when $\pi^{TT} = 0$, then (44) has the unique solution $N = -\frac{9}{4}\tau$ which is spatially constant.

We now show that even in the case that $\pi^{TT} = 0$, then H_{reduced} is still monotonically decreasing unless g is an Einstein metric (and hence hyperbolic). Indeed, if $\pi^{TT} = 0$, then $\frac{d^2}{dt^2}H_{\text{reduced}} = 0$, whereas the third derivative is given by

$$\frac{d^3}{dt^3}H_{\text{reduced}} = -6\tau^2 \int_M N(\partial_t\pi^{TT} \cdot \partial_t\pi^{TT})\mu_g^{-1} \leq 0 \tag{46}$$

with equality holding if and only if $\partial_t\pi^{TT} = 0$. Since $\pi^{TT} = 0$, $N = -\frac{9}{4}\tau$ is spatially constant. Substituting this result together with $\pi^{TT} = 0$ into the *ADM* evolution equation for π one finds that $\partial_t\pi^{TT} = 0$ if and only if the Ricci tensor $\mathrm{Ric}(g)$ of g satisfies

$$\mathrm{Ric}(g) = -\tfrac{2}{9}\tau^2 g \tag{47}$$

with sectional curvature $K(g) = -\frac{1}{9}\tau^2$ and scalar curvature $R(g) = -\frac{2}{3}\tau^2$. Since g is an Einstein space with negative Ricci curvature and since $n = 3$, g must be hyperbolic.

On the other hand, it is straightforward to verify that for fixed $\tau \in (-\infty, 0)$, the Cauchy data $(g, \pi) = (g, \frac{2}{3}\tau g^{-1}\mu_g)$, where g satisfies (47), is a solution of the Einstein constraint equations and thus provides Cauchy data for a vacuum spacetime on $\boldsymbol{R}^- \times M$, where M is a hyperbolic manifold. In fact, the resulting spacetime is expressible as

$$ds^2 = -\left(\frac{3}{\tau^2}\right)^2 d\tau^2 + \frac{3}{2\tau^2}\tilde{\gamma}_{ij}dx^i dx^j \tag{48}$$

where $\tilde{\gamma}$ is a fixed, i.e. τ independent, hyperbolic metric satisfying $R(\tilde{\gamma}) = -1$ and where we have for convenience used τ instead of $t = \frac{2}{3}\frac{1}{\tau^2}$ as time coordinate. Locally the metric (48) is identical to the $k = -1$ vacuum Robertson-Walker solution which is well-known to be flat. Globally, these spacetimes, one for each compact M of hyperbolic type, are obtainable as quotients of the interior of a future light cone in Minkowski by co-compact subgroups of the proper orthochronous Lorentz group which fixes that point and are sometimes known as Löbell spacetimes ([27]; see also [21], p. 136, for other remarks regarding these spacetimes). They each admit a global timelike homothety inherited from the homothetic Killing field $x^\mu \frac{\partial}{\partial x^\mu}$ of Minkowski space which is compatible with the quotients, but no other global Killing or conformal Killing symmetries ([11]).

It is worth remarking here that the foregoing arguments work equally well in lower ($n = 2$) and higher ($n \geq 4$) spatial dimensions but that the conclusions are somewhat less restrictive, although for different reasons. For $n = 2$, there is for any higher genus surface Σ_p, $p \geq 2$, an entire Teichmüller space of solutions obtainable by the above construction, whereas for $n \geq 4$, condition (47) does not imply constancy of the sectional curvature of g. Thus only for $n = 3$ is the solution to (47) necessarily unique on each hyperbolic manifold M by virtue of Mostow rigidity. In any dimension, however, the corresponding solutions, when they exist, are the only vacuum spacetimes for which $H_{\mathrm{reduced}} = |\tau|^n \int_M \mu_g$ is constant in time. Every other solution of Einstein's equations has H_{reduced} monotonically decreasing. In arbitrary dimension ($n \geq 2$) these special metrics, ${}^{(n+1)}g$, can be written explicitly as

$$ds^2 = -\left(\frac{n}{\tau^2}\right)^2 d\tau^2 + \left(\frac{n}{n-1}\right)\frac{1}{\tau^2}\tilde{\gamma}_{ij}dx^i dx^j \tag{49}$$

where $\tilde{\gamma}$ is a fixed, i.e., τ-independent, Einstein metric normalized by $R(\tilde{\gamma}) = -1$. When $n = 2$ or 3, or when $n \geq 4$ and $\tilde{\gamma}$ is hyperbolic, these metrics are flat and in every case admit a global homothetic Killing vector field $Y = \tau\frac{\partial}{\partial\tau}$ on the spacetime, with $L_Y {}^{(n+1)}g = -2\,{}^{(n+1)}g$, where L_Y denotes the Lie derivative with respect to Y.

Returning to the $n = 3$ case, we consider again the special Cauchy data $(g, \pi) = (g, \frac{2}{3}\tau g^{-1}\mu_g)$ discussed above. Note first that $\pi^{TT} = \varphi^{-4}p^{TT}$, so that $\pi^{TT} = 0$ if and only if $p^{TT} = 0$, in which case the unique positive solution of Lichnerowicz's equation is

$$\varphi = \left(\frac{3}{2\tau^2}\right)^{1/4} . \tag{50}$$

Thus the gravitational Cauchy data (g, π) corresponding to the special reduced Cauchy data $(\tau, \tilde{\gamma}, 0) \in \boldsymbol{R}^- \times P_{\mathrm{reduced}}$ is given by

$$\begin{aligned}(g, \pi) &= \left(g,\, \tfrac{2}{3}\tau g^{-1}\mu_g\right) = \left(\varphi^4\tilde{\gamma},\, \tfrac{2}{3}\tau(\varphi^4\tilde{\gamma})^{-1}\mu_{\varphi^4\tilde{\gamma}}\right) \\ &= \left(\varphi^4\tilde{\gamma},\, \tfrac{2}{3}\tau\varphi^2\tilde{\gamma}^{-1}\mu_{\tilde{\gamma}}\right) = \left(\tfrac{3}{2}\tau^{-2}\tilde{\gamma},\, -\left(\tfrac{2}{3}\right)^{1/2}\tilde{\gamma}^{-1}\mu_{\tilde{\gamma}}\right)\end{aligned} \tag{51}$$

where $\tilde{\gamma}$ is hyperbolic with $K(\tilde{\gamma}) = -\frac{1}{6}$.

We now show that for fixed τ the special reduced Cauchy data $(\tilde{\gamma}, 0)$ is, up to isometry, the only critical point of H_{reduced} and, furthermore, is a strict local minimum for H_{reduced}.

We first consider variations of $\int_M \mu_g = \int_M \varphi^6 \mu_\gamma$ with respect to the variable p^{TT} holding τ and γ fixed. Letting r^{TT} denote the variation of p^{TT} we compute

$$D\left(\int_M \varphi^6 \mu_\gamma\right) \cdot r^{TT} = 6\int_M \varphi^5 \Big(D\varphi(r^{TT})\Big)\mu_\gamma \ . \tag{52}$$

The variation $\delta\varphi = D\varphi \cdot (r^{TT})$ is determined by the corresponding variation of the Lichnerowicz equation (11), which gives

$$\begin{aligned} -\Delta_\gamma(\delta\varphi) + \tfrac{1}{8}(\delta\varphi) - \tfrac{5}{12}\tau^2\varphi^4(\delta\varphi) - \tfrac{7}{8}(p^{TT}\cdot p^{TT})\mu_\gamma^{-2}\varphi^{-8}(\delta\varphi) \\ + \tfrac{1}{4}(p^{TT}\cdot r^{TT})\mu_\gamma^{-2}\varphi^{-7} = 0 \end{aligned} \tag{53}$$

where $p^{TT} \cdot r^{TT} = (p^{TT})^{ij}(r^{TT})_{ij}$. Multiplying this equation by φ, integrating over M, and integrating by parts to arrange for the Laplacian to act on φ, we use the Lichnerowicz equation for φ to re-express this term and obtain

$$\tfrac{1}{3}\tau^2 \int_M \varphi^5(\delta\varphi)\mu_\gamma + \int_M (p^{TT}\cdot p^{TT})\varphi^{-7}\mu_\gamma^{-1}(\delta\varphi) = \tfrac{1}{4}\int_M (p^{TT}\cdot r^{TT})\mu_\gamma^{-1}\varphi^{-6} \ . \tag{54}$$

Clearly if $p^{TT} = 0$, we get $\int_M \varphi^5(\delta\varphi)\mu_\gamma = 0$ for all r^{TT}. To show that p^{TT} must vanish at a critical point, take $r^{TT} = p^{TT}$ and apply the maximum principle to (53) to show that at a minimum $(\delta\varphi)_{\min}$ for $\delta\varphi$ we have

$$\begin{aligned} &\left(\tfrac{5}{12}\tau^2\varphi^4 - \tfrac{1}{8} + \tfrac{7}{8}(p^{TT}\cdot p^{TT})\mu_\gamma^{-2}\varphi^{-8}\right)_{\text{min pt}} \delta\varphi_{\min} \\ &\geq \tfrac{1}{4}\Big((p^{TT}\cdot r^{TT})\mu_\gamma^{-2}\varphi^{-7}\Big)_{\text{min pt}} = \tfrac{1}{4}\Big((p^{TT}\cdot p^{TT})\mu_\gamma^{-2}\varphi^{-7}\Big)_{\text{min pt}} \geq 0 \ . \end{aligned} \tag{55}$$

The same maximum principle argument applied to the Lichnerowicz equation gives

$$\tfrac{1}{12}\tau^2\varphi_{\min}^4 - \tfrac{1}{8} \geq \tfrac{1}{8}\Big((p^{TT}\cdot p^{TT})\mu_\gamma^{-2}\Big)_{\text{min pt}}(\varphi_{\min}^{-8}) \ \geq 0 \tag{56}$$

and thus the quantity in parentheses on the left hand side of (55) is strictly positive, since φ is bounded away from zero. Thus (55) gives $(\delta\varphi)_{\min} \geq 0$ when $r^{TT} = p^{TT}$ and (54) implies that $\delta\varphi$ is not identically zero unless p^{TT} vanishes identically. Since $\varphi > 0$ everywhere it follows that $\int_M \varphi^5(\delta\varphi)\mu_\gamma > 0$ unless $p^{TT} = 0$ identically.

We conclude that $D\left(\int_M \varphi^6\mu_\gamma\right) \cdot r^{TT}$ vanishes for all r^{TT} if and only if p^{TT} vanishes identically. In this case, the unique positive solution of Lichnerowicz's equation $\varphi = \left(\frac{3}{2\tau^2}\right)^{1/4}$ is constant and thus independent of γ. Thus variations of φ with respect to γ vanish identically when $p^{TT} = 0$. Let h represent the variation of γ. Then

$$D\Big(-\tau^3\int \varphi^6\mu_\gamma\Big)\cdot h = (\tfrac{3}{2})^{3/2}\Big(D\int_M \mu_\gamma\Big)\cdot h \ . \tag{57}$$

Thus critical points of H_{reduced} *are those points with* $p^{TT} = 0$ *and* γ *a critical point of volume among metrics satisfying* $R(\gamma) = -1$.

Strictly speaking we should also restrict γ to lie in a slice for the $\mathcal{D}_0$ action on $\mathcal{M}_{-1}$ since the reduced configuration space is $\mathcal{M}_{-1}/\mathcal{D}_0$. However, since the volume functional $\int_M \mu_\gamma$ is invariant with respect to the $\mathcal{D}_0$-action (and indeed the full $\mathcal{D}$-action), its derivative in the fiber direction vanishes identically. For convenience therefore, we shall differentiate $\int_M \mu_\gamma$ subject only to the condition that the variation h be tangent to $\mathcal{M}_{-1}$.

Variations preserving $R(\gamma) = -1$ must obey $DR(\gamma)\cdot h = 0$ or, equivalently, for all $\sigma \in C^\infty(M, \boldsymbol{R})$,

$$\int_M \sigma\Big(DR(\gamma)\cdot h\Big)\mu_\gamma = \int_M h\cdot\Big(DR(\gamma)^*\cdot\sigma\Big)\mu_\gamma = 0 \ , \tag{58}$$

where $DR(\gamma)$ is given by

$$DR(\gamma)\cdot h = h^{ij}{}_{|i|j} - \gamma^{ij}(\gamma^{kl}h_{kl})_{|i|j} - R^{ij}h_{ij} \ , \tag{59}$$

and where, from an integration by parts, the L^2-dual $DR(\gamma)^*$ is given by

$$(DR(\gamma)^*\cdot\sigma)_{ij} = \sigma_{|i|j} - (\gamma^{kl}\sigma_{|k|l})\gamma_{ij} - \sigma R_{ij} \ . \tag{60}$$

Since the range of $DR(\gamma)^*$ is closed, $h \in \ker DR(\gamma)$ if and only if h is L^2-orthogonal to range $DR(\gamma)^*$ (see [12] for details). The critical point condition is thus

$$D\Big(\int_M \mu_\gamma\Big)\cdot h = \tfrac{1}{2}\int_M (\mathrm{tr}_\gamma h)\mu_\gamma = \tfrac{1}{2}\int_M (\gamma^{ij}h_{ij})\mu_\gamma = 0 \tag{61}$$

for all h orthogonal to range $DR(\gamma)^*$. But this means that γ must be parallel to range $DR(\gamma)^*$, i.e., that

$$\gamma_{ij} = \sigma_{|i|j} - \gamma_{ij}\sigma^{|k}{}_{|k} - \sigma R_{ij} \tag{62}$$

for some σ. Taking the trace of this equation and using $R(\gamma) = -1$ gives

$$3 = -2\sigma^{|k}{}_{|k} + \sigma \tag{63}$$

which has the unique solution $\sigma = 3$. From (62) it then follows that γ is an Einstein metric,

$$\mathrm{Ric}(\gamma) = -\tfrac{1}{3}\gamma \tag{64}$$

and since $n = 3$, γ must be hyperbolic, and thus by Mostow rigidity $\gamma \in \mathcal{M}_{-1}$ is unique up to isometry.

Thus we conclude that, firstly, $\gamma \in \mathcal{M}_{-1}$ *is a critical point of* vol *restricted to* $\mathcal{M}_{-1}$ *if and only if* γ *is hyperbolic* (with $K(\gamma) = -\frac{1}{6}$), and secondly, (γ, p^{TT}) *is a critical point of* H_{reduced} *if and only if* $p^{TT} = 0$ *and* γ *is a hyperbolic metric, in which case* $(\gamma, 0)$ *is a unique (up to isometry) isolated critical point of* H_{reduced}.

In (38), we have shown that if M is a compact connected oriented 3-manifold with $Y(M) = -1$, then $\sigma(M)$ is expressed in terms of the infimum of the volume of Riemannian metrics restricted to the submanifold $\mathcal{M}_{-1}$. Coupled with the result presented here, *this infimum can only be attained if γ is a hyperbolic metric.* We shall return to this result in Sects. 5 and 6.

We remark that when $n \geq 4$, the condition of being an Einstein metric does not imply hyperbolicity. Thus in these higher dimension cases, the critical points may not be isolated, even though Mostow rigidity still applies (but only to hyperbolic structures).

In the $n = 3$ case, we now show that when M is of hyperbolic type, then the unique isolated critical point $(\tilde{\gamma}, 0)$ is a strict local minimum. Here we sketch the proof of this result; technical details will appear elsewhere. The first step in the proof is accomplished by computing the Hessian of H_{reduced} at $(\tilde{\gamma}, 0)$ and verifying its positive definiteness there.

Let (h, r^{TT}) denote the variation of (γ, p^{TT}) about the critical point $(\tilde{\gamma}, 0)$ with $DR(\tilde{\gamma}) \cdot h = 0$ as above. We compute $D^2 H_{\text{reduced}} \cdot ((h, r^{TT}), (h, r^{TT})) = D^2 H_{\text{reduced}}(\tau, \tilde{\gamma}, 0) \cdot ((h, r^{TT}), (h, r^{TT}))$, suppressing the base point $(\tau, \tilde{\gamma}, 0)$,

$$\begin{aligned} &D^2\Big(-\tau^3 \int_M \varphi^6 \mu_{\tilde{\gamma}} \Big) \cdot \Big((h, r^{TT}), (h, r^{TT}) \Big) \\ &\quad = -30\tau^3 \int_M \varphi^4 \Big(D\varphi \cdot (h, r^{TT}) \Big)^2 \mu_{\tilde{\gamma}} - \tau^3 \int_M 6\varphi^5 D^2\varphi \Big((h, r^{TT}), (h, r^{TT}) \Big) \mu_{\tilde{\gamma}} \\ &\qquad -\tau^3 \int_M 12\varphi^5 \Big(D\varphi \cdot (h, r^{TT}) \Big) \Big(D\mu_{\tilde{\gamma}} \cdot h \Big) - \tau^3 \int_M \varphi^6 D^2 \mu_{\tilde{\gamma}} \cdot (h, h) \end{aligned} \tag{65}$$

where $D\mu_{\tilde{\gamma}} \cdot h = D\mu_{\tilde{\gamma}} \cdot (h, r^{TT})$ and $D^2 \mu_{\tilde{\gamma}} \cdot (h, h) = D^2 \mu_{\tilde{\gamma}} \cdot ((h, r^{TT}), (h, r^{TT}))$ since the volume functional μ_γ is a function of γ alone.

However, an arbitrary variation of (γ, p^{TT}) in the Lichnerowicz equation (11) simplifies greatly when $p^{TT} = 0$ since in that case $\varphi = (\frac{3}{2\tau^2})^{1/4}$ is constant so that $d\varphi = 0$, thereby yielding an equation for which the unique solution is $D\varphi \cdot (h, r^{TT}) = 0$. Thus the Hessian reduces to

$$\begin{aligned} &D^2 H_{\text{reduced}} \cdot \Big((h, r^{TT}), (h, r^{TT}) \Big) \\ &\quad = -6\tau^3 \int_M \varphi^5 D^2\varphi \Big((h, r^{TT}), (h, r^{TT}) \Big) \mu_\gamma - \tau^3 \int_M \varphi^6 D^2 \mu_\gamma \cdot (h, h) \end{aligned} \tag{66}$$

and one can compute the second variation $\delta^{(2)}\varphi = D^2\varphi \cdot ((h, r^{TT}), (h, r^{TT}))$ by solving the second variation of the Lichnerowicz equation which, at a critical point, reduces to

$$-\Delta_\gamma(\delta^{(2)}\varphi) + \tfrac{1}{8}(\delta^{(2)}\varphi) - \tfrac{5}{12}\tau^2\varphi^4(\delta^{(2)}\varphi) + \tfrac{1}{4}(r^{TT} \cdot r^{TT})\mu_\gamma^{-2}\varphi^{-7} = 0 \tag{67}$$

since the first variation $\delta\varphi$ vanishes identically there. Using $\varphi^4 = \frac{3}{2\tau^2}$ and integrating (67) over M, we have

$$-6\tau^3 \int_M \varphi^5 D^2\varphi \cdot \Big((h, r^{TT}), (h, r^{TT}) \Big) \mu_\gamma = 6^{1/2}\tau^4 \int_M (r^{TT} \cdot r^{TT}) \mu_\gamma^{-1} \ . \tag{68}$$

Thus we now need to evaluate the second term $-\tau^3 \int_M \varphi^6 D^2\mu_\gamma \cdot (h,h)$ in (66) for an arbitrary h which is tangent to the intersection of $\mathcal{M}_{-1}$ and the affine slice defined in [15]. Let $\gamma(\lambda)$ be a smooth curve lying in this intersection such that $\gamma(0) = \tilde{\gamma}$, the critical metric, $\gamma'(0) = h$, and let $\gamma''(0) = \ell$. Since by assumption h is tangent to $\mathcal{M}_{-1}$, we have

$$DR(\tilde{\gamma}) \cdot h = h^{ij}{}_{|i|j} - \tilde{\gamma}^{ij}(tr_{\tilde{\gamma}} h)_{|i|j} - \tilde{R}^{ij} h_{ij} = 0 \tag{69a}$$

$$(\delta_{\tilde{\gamma}} h)_i = -\tilde{\gamma}^{jk} h_{ij|k} = 0 \tag{69b}$$

where (69a) is the condition that h be tangent to $\mathcal{M}_{-1}$, (69b) is the condition that h be tangent to the affine slice, and where the divergence, covariant derivative, and Ricci tensor are with respect to $\tilde{\gamma}$. Combining these equations and using the fact that $\tilde{\gamma}$ is hyperbolic, one sees that $tr_{\tilde{\gamma}} h = 0$ and thus that h is transverse-traceless with respect to $\tilde{\gamma}$. We thus write h^{TT} for h in the following.

Since we are at a critical point for volume in $\mathcal{M}_{-1}$ we can compute the Hessian of volume by evaluating the second derivative of the volume along the curve $\gamma(\lambda)$ so that

$$\begin{aligned} -\tau^3 \int_M \varphi^6 D^2\mu_\gamma \cdot (h^{TT}, h^{TT}) &= (\tfrac{3}{2})^{3/2} \left(\frac{d^2}{d\lambda^2} \int_M \mu_{\gamma(\lambda)} \right)\Big|_{\lambda=0} \\ &= \tfrac{1}{2}(\tfrac{3}{2})^{3/2} \int_M \left(-h^{TT} \cdot h^{TT} + tr_{\tilde{\gamma}} \ell \right) \mu_{\tilde{\gamma}} \ . \end{aligned} \tag{70}$$

Now $\ell = \gamma''(0)$ is restricted by the condition that $\gamma(\lambda)$ lies in $\mathcal{M}_{-1}$ which implies that

$$0 = \frac{d^2}{d\lambda^2} R(\gamma(\lambda))\big|_{\lambda=0} = DR(\tilde{\gamma}) \cdot \ell + D^2R(\tilde{\gamma}) \cdot (h^{TT}, h^{TT}) \tag{71}$$

as well as a condition on ℓ which ensures that $\gamma(\lambda)$ remains in the slice through second order. This latter condition is simply that $\delta_{\tilde{\gamma}} \ell = 0$ which, together with (69a), (69b), and (71), yields

$$\tfrac{1}{3} tr_{\tilde{\gamma}} \ell - \tilde{\gamma}^{kl} (tr_{\tilde{\gamma}} \ell)_{|k|l} + D^2R(\tilde{\gamma}) \cdot (h^{TT}, h^{TT}) = 0 \ . \tag{72}$$

Integrating this equation over M gives the needed formula for $\int_M (tr_{\tilde{\gamma}} \ell) \mu_{\tilde{\gamma}}$ in terms of h^{TT}, namely

$$\int_M (tr_{\tilde{\gamma}} \ell) \mu_{\tilde{\gamma}} = -3 \int_M D^2R(\tilde{\gamma}) \cdot (h^{TT}, h^{TT}) \mu_{\tilde{\gamma}} \tag{73}$$

which one can evaluate using standard second variation results for the scalar curvature. The final result is

$$\begin{aligned} \frac{d^2}{d\lambda^2} \left(\int_M \mu_{\gamma(\lambda)} \right)\Big|_{\lambda=0} &= \tfrac{3}{4} \int_M \left(\nabla_{\tilde{\gamma}} h^{TT} \cdot \nabla_{\tilde{\gamma}} h^{TT} - \tfrac{1}{3} h^{TT} \cdot h^{TT} \right) \mu_{\tilde{\gamma}} \\ &= \tfrac{3}{4} \int_M \left(\|\nabla_{\tilde{\gamma}} h^{TT}\|^2_{\tilde{\gamma}} - \tfrac{1}{3} \|h^{TT}\|^2_{\tilde{\gamma}} \right) \mu_{\tilde{\gamma}} \end{aligned} \tag{74}$$

(in local coordinates, the integrand is given by $\tilde{\gamma}^{ij}\tilde{\gamma}^{kl}\tilde{\gamma}^{mn}(h^{TT})_{ik|m}(h^{TT})_{jl|n} - \frac{1}{3}\tilde{\gamma}^{ij}\tilde{\gamma}^{kl}(h^{TT})_{ik}(h^{TT})_{jl}$). From (74) we can conclude our calculation of the Hessian (66) of H_{reduced} to give

$$D^2 H_{\mathrm{reduced}}(\tau,\tilde{\gamma},0)\cdot\Big((h^{TT},r^{TT}),(h^{TT},r^{TT})\Big) \tag{75}$$
$$= 6^{1/2}\tau^4\int_M \left(r^{TT}\cdot r^{TT}\right)\mu_{\tilde{\gamma}}^{-1} + \tfrac{3}{4}(\tfrac{3}{2})^{1/2}\int_M \Big(\|\nabla_{\tilde{\gamma}}h^{TT}\|^2_{\tilde{\gamma}} - \tfrac{1}{3}\|h^{TT}\|^2_{\tilde{\gamma}}\Big)\mu_{\tilde{\gamma}}\ .$$

To show that this is positive definite, in spite of the final negative term, we need the following identity for arbitrary transverse-traceless tensors on a compact hyperbolic 3-manifold $(M,\tilde{\gamma})$ with sectional curvature $K(\tilde{\gamma}) = -\frac{1}{6}$,

$$\tfrac{1}{2}\int_M \Big(h^{TT}_{jm|l} - h^{TT}_{lm|j}\Big)\Big((h^{TT})^{jm|l} - (h^{TT})^{lm|j}\Big)\mu_{\tilde{\gamma}}$$
$$= \int_M \Big((h^{TT})_{jm|l}(h^{TT})^{jm|l} - \tfrac{1}{2}(h^{TT})_{jm}(h^{TT})^{jm}\Big)\mu_{\tilde{\gamma}}$$
$$= \int_M \Big(\|\nabla_{\tilde{\gamma}}h^{TT}\|^2_{\tilde{\gamma}} - \tfrac{1}{2}\|h^{TT}\|^2_{\tilde{\gamma}}\Big)\mu_{\tilde{\gamma}} \tag{76}$$

which is obtained by integrating by parts with respect to the hyperbolic background metric $\tilde{\gamma}$. We thank Robert Beig for providing us with this identity (see also [8]). Comparing (75) with (76) shows that the "potential" term (76) is positive definite. With further technical details, we can then conclude that $(\tilde{\gamma},0)$ is a strict local minimum of H_{reduced}.

Thus, if M is a hyperbolic manifold and τ is fixed, the unique (up to isometry) critical point $(\tilde{\gamma},0) \in P_{\mathrm{reduced}}$ of $H_{\mathrm{reduced}}(\tau,\gamma,p^{TT})$ is a strict local minimum (in the non-isometric direction).

5 The Infimum of H_{reduced} and Its Relation to the σ-Constant of M

We have shown that if M is a hyperbolic manifold, then for each fixed τ, H_{reduced} has a a unique (up to isometry) critical point $(\tilde{\gamma},0) \in P_{\mathrm{reduced}}$ which is a strict local minimum of H_{reduced}. It is tempting to speculate that this local minimum is in fact a strict global minimum.

To put this conjecture into a larger context let us return to an arbitrary M of Yamabe type -1 and study the infimum of $H_{\mathrm{reduced}}(\tau,\gamma,p^{TT})$.

We first show that along each fiber above a fixed conformal metric γ, H_{reduced} is minimized by setting the fiber variable $p^{TT} = 0$. This will reduce the study of the infimum of H_{reduced} to the study of the infimum of the volume functional on $\mathcal{M}_{-1}$. But the volume of a metric lying in $\mathcal{M}_{-1}$ is directly related to the Yamabe invariant of its associated conformal class. Taking the supremum of the Yamabe invariant over all conformal classes parameterized by $\mathcal{M}_{-1}$ corresponds to taking the infimum of volume on $\mathcal{M}_{-1}$ (see Sect. 2.4), thus leading to a relation between the σ-constant of M and the infimum of H_{reduced} on the phase space P_{reduced}.

To see how H_{reduced} depends upon the p^{TT} variable one can investigate its variation along a "ray" $p^{TT}(\lambda) = \lambda p^{TT}$. Thus, fixing τ and γ, by a computation completely analogous to that which led to Eqn (55), one shows that the first order variation $\delta\varphi = \frac{d\varphi}{d\lambda}(\tau, \gamma, \lambda p^{TT})$ is strictly positive for $\lambda > 0$ unless p^{TT} is identically zero. Thus $\frac{d}{d\lambda}(-\tau^3 \int_M \varphi^6 \mu_\gamma) \geq 0$ for $\lambda > 0$ with equality holding only if $p^{TT} = 0$. Conversely, it follows that we can scale H_{reduced} to its minimum along each fiber by scaling the fiber variable to zero. Thus H_{reduced} is minimized at fixed (τ, γ) by putting $p^{TT} = 0$ for which φ becomes $(\frac{3}{2\tau^2})^{1/4}$ and then

$$H_{\text{reduced}}(\tau, \gamma, 0) = -\tau^3 \int_M \varphi^6 \mu_\gamma = (\tfrac{3}{2})^{3/2} \int_M \mu_\gamma \quad . \tag{77}$$

Continuing to keep τ fixed but letting $\gamma \in \mathcal{M}_{-1}$ vary, we find

$$\begin{aligned} \inf_{(\gamma, p^{TT}) \in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, p^{TT})) &= \inf_{(\gamma, 0) \in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, 0) \\ &= \inf_{\gamma \in \mathcal{M}_{-1}} H_{\text{reduced}}(\tau, \gamma, 0) \\ &= (\tfrac{3}{2})^{3/2} \inf_{\gamma \in \mathcal{M}_{-1}} \left(\int_M \mu_\gamma \right) \\ &= (\tfrac{3}{2})^{3/2} \inf_{\gamma \in \mathcal{M}_{-1}} \Big(\text{vol}(M, \gamma) \Big) \quad . \end{aligned} \tag{78}$$

Thus the infimum of H_{reduced} is $(\frac{3}{2})^{3/2}$ times the infimum of volume restricted to $\mathcal{M}_{-1}$. In particular, if the infimum of H_{reduced} is achieved by a $\gamma \in \mathcal{M}_{-1}$, then γ must be hyperbolic.

Thus, combining (38) and (78), we find that for fixed $\tau \in (-\infty, 0)$ and M of Yamabe type -1, *then the σ-constant is related to the infimum of the reduced Hamiltonian as follows:*

$$\sigma(M) = -\tfrac{2}{3} \left(\inf_{(\gamma, p^{TT}) \in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, p^{TT}) \right)^{2/3} \tag{79}$$

or, alternately,

$$\inf_{(\gamma, p^{TT}) \in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, p^{TT}) = \Big(-\tfrac{3}{2}\sigma(M) \Big)^{3/2} \quad . \tag{80}$$

6 Concluding Remarks

For 3-manifolds of Yamabe type -1, we have seen in Sect. 4 that the σ-constant can only be achieved, if at all, by a metric having constant curvature and it has been conjectured that for a 3-manifold of hyperbolic type, the

Yamabe invariant for the hyperbolic metric $\tilde{\gamma}$ with $R(\tilde{\gamma}) = -1$ in fact equals the σ-constant for that manifold, $y(\tilde{\gamma}) = \sigma(M)$. If so, then our foregoing analysis would show that on a manifold of hyperbolic type, the global minimum of H_{reduced} is attained at $(\tilde{\gamma}, 0)$. As discussed in Sect. 4, this is precisely the Cauchy data for a certain canonical flat spacetime on $\boldsymbol{R}^{-} \times M$ which admits a global timelike homothety.

Such a result, if true, would be somewhat analogous to that which shows for asymptotically flat spacetimes that the global minimum of the *ADM*-mass functional is attained by the Cauchy data for flat Minkowski space. For now, however, we only know that the hyperbolic data described above yields a strict local minimum that may or may not be a strict global minimum.

The universal monotonic decay of H_{reduced}, together with the fact that the infimum of H_{reduced} determines the σ-constant of M, suggests that if a metric exists which realizes the σ-constant, then this metric is an attractor for the Einstein flow.

On the other hand, if M is of non-hyperbolic type, then no metric can realize $\sigma(M)$. In this case the Einstein flow is nevertheless seeking to attain the σ-constant asymptotically insofar as the reduced Hamiltonian is monotonically seeking to decay to its infimum. There may well be obstructions, such as the formation of black holes, which prevent any particular solution from approaching the σ-constant asymptotically but it seems plausible that some subset of solutions might nevertheless asymptote to this ideal attractor.

In the case that M is of hyperbolic type, some results of [3] show that in certain cases solutions sufficiently near the canonical flat solution as given by (48) do indeed asymptotically decay in the direction of expansion to the hyperbolic geometry. These cases occur for those hyperbolic 3-manifolds which do not admit non-trivial traceless Codazzi tensors, where a symmetric 2–tensor h is a called a *Codazzi tensor* if

$$h_{ij|k} - h_{ik|j} = 0 \quad . \tag{81}$$

Thus we note that a traceless Codazzi tensor is automatically transverse (divergence-free).

If there do not exist non-trivial traceless Codazzi tensors, then there is no non-trivial moduli space of flat spacetime perturbations and the 4-dimensional geometries for these can be entirely controlled by the higher order Bel-Robinson energies defined and studied in [3]. Thus the hyperbolic metric on such manifolds is indeed an attractor for the Einstein flow, and if the conjecture on $\sigma(M)$ is true in these cases, then this flow actually attains the σ-constant asymptotically.

Certain other hyperbolic 3-manifolds, namely, those which do admit non-trivial traceless Codazzi tensors, are not rigid in this 4-dimensional sense and admit a moduli space of flat perturbations of the canonical flat metric even though Mostow rigidity prohibits a deformation of the hyperbolic structure itself. For such flat perturbations the Bel-Robinson energies, based as they are entirely on curvature, cannot control the spacetime geometries and one needs an additional tool for the proof of long-time existence and asymptotic behavior. The reduced Hamiltonian may provide precisely the needed tool since it has an isolated local minimum at the hyperbolic data and decays towards this minimum in the direction of expansion. This Hamiltonian bounds at best an $H^1 \times L^2$ Sobolev norm of the Cauchy data (g, π) which is far too weak to use for the desired long-time-existence theorem. But all we really need is an additional bound on the finite dimensional space of modular parameters to complement the Bel-Robinson bounds on curvature and this, it seems, H_{reduced} can provide. Thus a modest but useful potential application of our results would be to use them to complete a proof of long-time-existence and asymptotic behavior for the case of non-rigid hyperbolic M.

In this context, it is worth remarking that in their study of the non-linear stability of Minkowski space, Christodoulou and Klainerman [10] never needed to appeal to the *ADM*-mass functional, which is roughly the analogue of H_{reduced} here, but only used Bel-Robinson type energies. However, Minkowski space is known to be isolated as a flat solution of Einstein's equations for the case of asymptotically flat spacetimes and this is more analogous to one of our rigid cases which also seem only to require curvature type energy estimates for their analysis.

Perhaps the most interesting potential application of our results is to the case of a manifold M of non-hyperbolic-type for which the σ-constant can never be realized by an actual metric but only approached as a limit through a sequence of metrics. Anderson ([1], [2]) has formulated a set of conjectures about how a sequence of metrics degenerates when the σ-constant is approached which, if true, would imply the Thurston conjectures [41] and thus complete the 3-manifold classification program. But if the Einstein flow is seeking to attain $\sigma(M)$ asymptotically for M of non-hyperbolic-type, then presumably the curve of conformal geometries defined by any particular solution for which H_{reduced} tends to its infimum is degenerating in precisely the way outlined by Anderson's conjectures.

If this is the case then perhaps the asymptotic behavior of large classes of Einstein spacetimes can be characterized rather explicitly in terms of the degenerations described in Anderson's conjectures. Conversely, it is not inconceivable that the Einstein flow, much like the Ricci flow before it ([20]), could be used in a positive way to try to establish some form of the geometrization conjectures for 3-manifolds.

A model for many of the above concepts is provided by $2+1$ dimensional gravity defined on manifolds of the form $\boldsymbol{R}^- \times \Sigma_p$, $p \geq 2$. In the $2+1$ case, restricting to surfaces with $p \geq 2$ is the analogue in the $3+1$ case of our restriction to manifolds of Yamabe type -1. In $2+1$ gravity the role of the σ-constant is played by the Euler characteristic $\chi(\Sigma_p)$ of the surface Σ_p. In fact, identifying $\tau^2 \int_{\Sigma_p} \mu_g$ with the reduced Hamiltonian, one derives an expression for H_{reduced} by integrating the Hamiltonian constraint over Σ_p

and appealing to the Gauss-Bonnet theorem to express the integral of the scalar curvature of (Σ_p, g) in terms of $\chi(\Sigma_p)$. The result is then expressed in terms of the reduced phase space variables (τ, γ, p^{TT}), as in [29], by setting

$$(g, \pi) = (e^{2\lambda}\gamma,\, e^{-2\lambda}p^{TT} + \tfrac{1}{2}\tau\gamma^{-1}\mu_\gamma) \tag{82}$$

where $R(\gamma) = -1$ and λ is the unique solution of the corresponding Lichnerowicz equation

$$\Delta_\gamma\lambda - \tfrac{1}{2} + \tfrac{1}{4}\tau^2 e^{2\lambda} - \tfrac{1}{2}e^{-2\lambda}(p^{TT}\cdot p^{TT})\mu_\gamma^{-2} = 0 \ . \tag{83}$$

One obtains

$$\begin{aligned} H_{\text{reduced}}(\tau, \gamma, p^{TT}) &= \tau^2 \int_{\Sigma_p} e^{2\lambda}\mu_\gamma = \tau^2 \int_{\Sigma_p} \mu_g \\ &= 2\int_{\Sigma_p} \Big((p^{TT}\cdot p^{TT})\mu_g^{-1} - R(g)\mu_g\Big) \\ &= 2\int_{\Sigma_p} (p^{TT}\cdot p^{TT})\mu_g^{-1} - 8\pi\chi(\Sigma_p) \ . \end{aligned} \tag{84}$$

It follows at once that the infimum of H_{reduced} is attained by setting $p^{TT} = 0$, a condition which is necessarily conserved by Einstein's equations since H_{reduced} can only decrease or remain constant at its minimum value which, in this case, is $-8\pi\chi(\Sigma)$. Thus for τ fixed, $p \geq 2$,

$$\inf_{(\gamma, p^{TT})\in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, p^{TT}) = -8\pi\chi(\Sigma_p) = 16\pi(p-1) > 0 \tag{85}$$

where

$$\begin{aligned} P_{\text{reduced}} = \{(\gamma, p^{TT}) \mid \gamma \in \mathcal{M}_{-1}(\Sigma_p) \ \text{and} \\ p^{TT} \ \text{is transverse-traceless with respect to}\ \gamma\,\} \end{aligned}$$

is the reduced phase space for Σ_p. This result may be thought of as the $2+1$ analogue of (80)

$$\inf_{(\gamma, p^{TT})\in P_{\text{reduced}}} H_{\text{reduced}}(\tau, \gamma, p^{TT}) = \left(\tfrac{3}{2}\right)^{3/2}\Big(-\sigma(M)\Big)^{3/2}$$

for $3+1$ gravity, showing again how in some sense the σ-constant for $n \geq 3$ acts as a generalization of the Euler characteristic for $n = 2$ (see Sect. 2.4).

An important difference between the cases of $2+1$ versus $3+1$ is that in the $2+1$ case, the particular solutions determined by setting $p^{TT} = 0$ are not isolated but are parameterized by the points of the Teichmüller space $\mathcal{T}(\Sigma_p)$ of Σ_p which, in turn, can be represented by metrics of constant negative curvature on Σ_p. These special solutions each admit a global timelike homothety as discussed above in Sect. 4 and it is known that they define an attractor for all the remaining solutions of vacuum $2+1$ dimensional gravity in the sense that every solution decays asymptotically to one of these special families in the limit of infinite expansion; details of this will appear elsewhere [4]. Unlike the case of hyperbolic 3-manifolds, where the attractor is an isolated fixed point, the space of such attractors has positive dimension $= \dim \mathcal{T}(\Sigma_p) = 6p - 6$, $p \geq 2$, since for $n = 2$, Mostow rigidity does not apply.

As we have mentioned, the case of $n+1$ dimensional gravity is potentially interesting for $n \geq 4$ since then the space of critical points of H_{reduced} is expected to coincide with the moduli space of Einstein metrics on M. We did not quite show this above but we did show that these are the only points at which H_{reduced} remains constant. It would be interesting to investigate whether such "critical points" are local, or conceivably global, minima of H_{reduced} and to see what light this question sheds on the topology of M.

Acknowledgements

The authors would like to thank Professors Robert Beig and Peter Aichelburg for cordially inviting us to the *Erwin Schrödinger Institute* in Vienna during the spring of 1997 where part of this research was carried out and to acknowledge the gracious hospitality and the financial support of the *ESI* during this period. Vincent Moncrief is also grateful to the *ESI* for their hospitality and support during the summer and fall of 1998 when part of this research was completed and to the *Institute for Theoretical Physics* at the University of California, Santa Barbara, for their hospitality and support during the spring of 1999 when this paper was completed.

The authors would also like to thank Professors Jürgen Ehlers, Helmut Friedrich, and Bernt Schmidt for cordially inviting us to the *Albert Einstein Institute* in Potsdam during the summer of 1997 during which time part of this research was carried out and to acknowledge the gracious hospitality and financial support of the *AEI* during this period.

The authors would also like to acknowledge informative discussions with Professors Lars Andersson, Robert Beig, and Alan Rendall.

This research was supported in part by NSF grants PHY-9503133 and PHY-9732629 to Yale University.

References

1. Anderson, M (1993), *Degeneration of metrics with bounded curvature and applications to critical metrics of Riemannian functionals*, Proceedings of Symposia in Pure Mathematics **54**, Part 3, 53–79.
2. Anderson, M (1997), *Scalar curvature and geometrization conjectures for* 3-*manifolds*, in *Comparison geometry* (Berkeley, CA 1993-94), Math. Sci. Res. Inst. Publ. **30**, Cambridge University Press, Cambridge.

3. Andersson, L, and Moncrief, V, *The global existence problem in general relativity*, to appear.
4. Andersson, L, and Moncrief, V, *Asymptotic behavior of solutions in* 2+1 *gravity*, in preparation.
5. Andersson, L, Moncrief, V, and Tromba, A (1997), *On the global evolution problem in* $(2+1)$*-gravity*, J. Geom. Phys. **23**, 191–205.
6. Arnowitt, R, Deser, S, and Misner, C (1962), *The dynamics of general relativity*, in *Gravitation: an introduction to current research*, edited by L Witten, John Wiley and Sons, Inc., New York.
7. Aubin, T (1982), *Nonlinear Analysis on Manifolds. Monge-Ampère Equations*, Springer-Verlag, New York.
8. Beig, R (1997), *TT-tensors and conformally flat structures on* 3*-manifolds*, in *Mathematics of Gravitation, Part 1, Lorentzian geometry and Einstein equations*, edited by P Chrusciel, Banach Center Publications **41**, 109–118.
9. Choquet-Bruhat, Y, and York, J (1980), *The Cauchy Problem*, in *General Relativity and Gravitation: Volume 1*, edited by A Held, Plenum Press, New York.
10. Christodoulou, D, and Klainerman, S (1993), *The global nonlinear stability of the Minkowski space*, Princeton University Press, Princeton, New Jersey.
11. Eardley, D, Isenberg, J, Marsden, J, and Moncrief, V (1986), *Homothetic and conformal symmetries of solutions to Einstein's equations*, Commun. Math. Phys. **106**, 137-158.
12. Fischer, A, and Marsden, J (1975), *Deformations of the scalar curvature*, Duke Mathematical Journal **42**, 519–547.
13. Fischer, A, and Moncrief, V (1994), *Reducing Einstein's equations to an unconstrained Hamiltonian system on the cotangent bundle of Teichmüller space*, in *Physics on Manifolds, Proceedings on the International Colloquium in honour of Yvonne Choquet-Bruhat*, edited by M Flato, R Kerner, and A Lichnerowicz, Kluwer Academic Publishers, Boston, 111–151.
14. Fischer, A, and Moncrief, V (1996), *The structure of quantum conformal superspace*, in *Global Structure and Evolution in General Relativity*, edited by S Cotsakis and G Gibbons, Springer-Verlag, Berlin, 111–173.
15. Fischer, A, and Moncrief, V (1997), *Hamiltonian reduction of Einstein's equations of general relativity*, Nuclear Physics B (Proc. Suppl.) **57**, 142–161.
16. Fischer, A, and Tromba, A (1984), *On a purely "Riemannian" proof of the structure and dimension of the unramified moduli space of a compact Riemann surface*, Mathematische Annalen **267**, 311–345.
17. Gromov, M, and Lawson, Jr., H (1980), *Spin and scalar curvature in the presence of a fundamental group. I*, Annals of Mathematics, **111**, 209–230.
18. Gromov, M, and Lawson, Jr., H (1980), *The classification of simply connected manifolds of positive scalar curvature*, Annals of Mathematics, **111**, 423–434.
19. Gromov, M, and Lawson, Jr., H (1983), *Positive scalar curvature and the Dirac operator on complete Riemannian manifolds*, Institut des Hautes Études Scientifiques, Publications Mathématiques, Number **58**, 83–196.
20. Hamilton, R (1982), *Three manifolds with positive Ricci curvature*, Jour. Diff. Geometry **17**, 255–306.
21. Hawking, S, and Ellis, G (1973), *The Large Scale Structure of Space-Time*, Cambridge University Press, Cambridge, England.
22. Kneser, H (1929), *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jahresbericht der Deutschen Mathematiker Vereinigung **38**, 248–260.

23. Kobayashi, S, and Nomizu, K (1963), *Foundations of Differential Geometry*, vol 1, Interscience, Wiley, New York.
24. Lanczos, C (1966), *The Variational Principles of Mechanics*, third edition, University of Toronto Press, Toronto.
25. Lawson, Jr., H, and Michelsohn, M-L (1989), *Spin Geometry*, Princeton University Press, Princeton, New Jersey.
26. Lee, J, and Parker, T (1987), *The Yamabe problem*, Bull. Amer. Math. Soc. **17**, 37–91.
27. Löbell, F (1931), *Beispiele geschlossener drei-dimensionaler Clifford-Kleinscher Räume negativer Krümmung*, Ber. Verhandl. Sächs. Akad. Wiss. Leipzig, Math. Phys. Kl. **83**, 167–174.
28. Milnor, J (1962), *A unique decomposition theorem for* 3*-manifolds*, American Journal of Mathematics **84**, 1–7.
29. Moncrief, V (1989), *Reduction of the Einstein equations in* $2+1$ *dimensions to a Hamiltonian system over Teichmüller space*, J. Math. Phys. **30** (12), 2907–2914.
30. Moncrief, V (1990), *How solvable is* $(2+1)$*-dimensional Einstein gravity?*, J. Math. Phys. **31** (12), 2978–2982.
31. Palais, R (1970), *Critical point theory and the minimax principle*, in *Global Analysis, Proceedings of Symposia in Pure Mathematics of the American Mathematical Society,* **15**, 185-212.
32. Palais, R, and Terng, C (1988), *Critical point theory and submanifold geometry*, Lecture Notes in Mathematics, **1353**, Springer-Verlag, New York.
33. Rabinowitz, P (1986), *Minimax methods in critical point theory with applications to differential equations*, Conference Board of the Mathematical Sciences, Number 65, American Mathematical Society, Providence, Rhode Island.
34. Schoen, R (1984) *Conformal deformation of a Riemannian metric to constant scalar curvature*, Jour. Diff. Geometry **20**, 479–495.
35. Schoen, R (1991), *On the number of constant scalar curvature metrics in a conformal class*, in *Differential geometry, a symposium in honour of Manfredo do Carmo*, edited by B Lawson and K Tenenblat, Longman Scientific and Technical, published in the United States with John Wiley, New York.
36. Schoen, R, and Yau, S-T (1979), *Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature*, Annals of Mathematics, **110**, 127–142.
37. Schoen, R, and Yau, S-T (1979), *The structure of manifolds with positive scalar curvature*, Manuscripta Math., **28**, 159–183.
38. Schoen, R, and Yau, S-T (1988), *Conformally flat manifolds, Kleinian groups and scalar curvature*, Invent. Math., **92**, 47–71.
39. Schoen, R, and Yau, S-T (1994), *Lectures on differential geometry*, Volume I, International Press, Cambridge, Massachusetts.
40. Scott, P (1983), *The geometries of* 3*-manifolds*, Bulletin London Math. Soc. **15**, 401-487.
41. Thurston, W (1997), *Three-dimensional geometry and topology*, volume 1, edited by S Levy, Princeton University Press, Princeton, New Jersey.
42. Wolf, J (1972), *Spaces of Constant Curvature*, second edition, Publish or Perish Press, Berkeley, California.

Anti-de-Sitter Spacetime and Its Uses

G. W. Gibbons

D.A.M.T.P., Cambridge University,
Silver Street, Cambridge CB3 9EW, U.K.

Abstract. This is a pedagogical account of some of the global properties of Anti-de-Sitter spacetime with a view to their application to the AdS/CFT correspondence. Particular care is taken over the distinction between Anti-de-Sitter and its covering space. It is argued that it is the former which is important.

1 Introduction

Because it is among the simplest of curved spacetimes, n-dimensional Anti-de-Sitter spacetime (AdS) has been of continuing interest to relativists. It has, since the earliest times of our subject, provided a test bed and a source of simple examples on which to try out novel ideas and spacetime concepts, both classical and quantum. It is a remarkable feature of the current search for a reformulation of the entire basis of theoretical physics, often referred to as M-theory, that of many of those older speculations find a natural home in, and have relevance for, present day efforts. This point will be amply illustrated in what follows.

Because it is homogeneous and has a large isometry group, $SO(n-1,2)$, AdS_n is the natural arena for enquiring to what extent the (Wignerian) group-theoretic ideas underlying relativistic quantum mechanics and quantum field theory in Minkowski spacetime $\mathbb{E}^{n-1,1}$, with isometry group the Poincaré group $E(n-1,1)$, extend to other spacetimes. Similar remarks apply to ideas about energy momentum and angular momentum conservation. The definitions of the ADM mass in General Relativity and the question of its positivity, which are closely linked, via Noether's theorem, to the properties of the isometry group [48], .

When quantising field theories we often seek a background or "ground state" around which to perform a perturbation expansion and AdS_n together with de-Sitter spacetime, dS_n with isometry group $SO(n,1)$, and Minkowski spacetime exhaust the list of maximally symmetric ground states. While de-Sitter spacetime arises naturally in studies of inflation, Anti-de-Sitter spacetime arises as the natural ground state of gauged supergravity theories.

We can regard flat space as a limit of the de-Sitter spacetimes as the cosmological constant goes to zero. In the process the isometry group undergoes a Wigner-Inönü [2] contraction to the Poincaré group. It is interesting to note therefore that a simple Lie-algebra cohomology argument gives a converse: these are the only isometry groups that may be obtained in this way [3].

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 102–142, 2000.

A major topic of interest in quantum gravity is the extent to which the global and topological properties of spacetime, such as the existence of closed timelike curves (CTCs), spatial compactness etc, feed into the quantum theory. Indeed there is a more basic question: how do geometrical and spacetime concepts themselves translate into quantum mechanical language? In the case of de-Sitter and Anti-de-Sitter spacetimes, with space and time topology $S^{n-1} \times \mathbb{R}$ and $\mathbb{R}^{n-1} \times S^1$ respectively, and because of the high degree of symmetry, these questions may frequently be translated into group-theoretic language which may then admit a simple group-theoretic answer. In this connection it is essential to be aware of the many important differences between the properties of the compact Lie groups with which particle physicists are most often familar and those of the isometry groups of Lorentzian spacetimes which are almost always noncompact[1].

Currently a great deal of attention has been focussed on Anti-de-Sitter spacetimes because (multiplied by a sphere) they arise as the near horizon geometry of the extreme black holes and extreme p-branes which play such an an essential role in our understanding of M-theory. This has led to Maldacena's AdS/CFT correspondence conjecture which places AdS and indeed Euclidean quantum gravity at the centre stage. In an interesting parallel and closely linked development, the mass and event horizon area properties of topologically nontrivial black holes, which can only arise in Anti-de-Sitter backgrounds have also attracted a great deal of interest recently.

In the notes which follow, I shall argue that it is fruitful if not essential to view these recent problems, like the former ones, with the correct global perspective and that if one does so one arrives at what at first may appear to be some surprising and counter-intuitive conclusions. For that reason, and in view of the audience's interests, I shall be concentrating on the basic geometrical and group theoretic descriptions rather on the more technical details concerning supersymetry, supergravity and superstring theory. For an earlier account with more emphasis on the supergravity applications the reader is referred to [5]. One striking feature, which is especially appropriate for this meeting, is that much of the discussion can be couched in the simple geometrical terms which would have been accessible to scientific workers in this city, and possibly on this very spot, two and a half millennia ago.

2 M-Theory

By way of motivation, recall that whatever it finally turns out to be, M-theory is a theory about *p-branes*, that is extended objects with p spatial dimensions moving in some higher dimensional spacetime, usually eleven dimensions. Thus $p = 0$ are point particles, $p = 1$ are strings $p = 2$ are membranes etc. The case $p = -1$ arises as "instantons".

[1] Lorentzian Taub-NUT spacetimes with isometry group $SU(2)$ or $SO(3)$ are an interesting exception

2.1 Levels of Description

Currently we have various levels of description at various levels of approximation for dealing with branes in M-theory.

- As D-branes, that is as the end points of fundamental or F-strings subject to Dirichlet boundary conditions. At this level it is believed that one may use the techniques of two-dimensional conformal field theory (CFT) to give a fully quantum mechanical treatment.
- As "soliton" solutions of classical supergravity theories. This is the "heavy" brane approximation which takes into account their self-gravity and is believed to be applicable in the semiclassical approximation when a large number, N, of light branes sit on top of one another. The solutions one starts with are typically static, have extreme Killing horizons and are BPS, which means that they admit some Killing spinor fields of the associated supergravity theory.
- As classical solutions of a Dirac-Born-Infeld Lagrangian describing a "light" brane, thought of as a $(p+1)$-dimensional timelike submanifold Σ_{p+1} moving in a fixed spacetime background M. The equations of motion are a generalisation of the standard equations for a minimal submanifold because in addition to the embedding map $x : \Sigma_{p+1} \to M$ (which provides scalar fields on the world volume Σ_{p+1}) each D-brane carries an abelian gauge field A_μ which may be viewed as $U(1)$ connection on a bundle over Σ_{p+1}. From the string theory standpoint, this vector field is associated with a open string of almost vanishing length, beginning and ending on the D-brane. Because the string has almost vanishing length it has almost vanishing energy and gives rise to a "light state" associated with the massless gauge field A_μ.

Strictly speaking the list given above does not exhaust all current brane descriptions because it omits the M5-brane action. However the details of the M5-brane action will not play an essential role in the future discussion.

2.2 Symmetry Enhancement

If one has N branes, one has N $U(1)$ gauge fields. Now as the branes coalesce one might have supposed one would get a description in which one has a $U(1)^N$ gauge theory over the coalesced brane world volume $\bar{\Sigma}_{p+1}$. However, from the string standpoint it is clear that $N(N-1)$ extra "light states" appear associated with strings of almost vanishing length beginning on one of the N strings and ending on another. This gives rise to a total of N^2 massless gauge fields on $\bar{\Sigma}_{p+1}$. Again one might have supposed that this would give rise to a description in which one has a $U(1)^{N^2}$ gauge theory on $\bar{\Sigma}_{p+1}$. However, in a way which so far has only been understood in detail using conformal field theory, a process of nonabelian symmetry enhancement is believed to occur and the resultant gauge group becomes nonabelian, and in fact $U(N)$. The $U(1)$ factor is associated to the centre of mass motion of the D-brane.

2.3 Killing Spinors

A supersymmetric solution of a supergravity theory is a solution admitting one or more spinor fields ϵ satisfying

$$\nabla\epsilon + N\epsilon = 0, \tag{1}$$

where ∇ is the Levi-Civita connection and N is a Clifford algebra valued one-form. The form of N depends on the details of the supergravity theory concerned. If $N = 0$, then a Killing spinor must be covariantly constant. This leads to the study of those holonomy groups which stabilise a spinor. The examples best known to relativists are the pp-waves. In AdS_n one has

$$N_\alpha = \pm\frac{1}{2R}\gamma_\alpha \tag{2}$$

with $\alpha = 0, 1 \ldots, n-1$. One easily verifies that, for either choice of sign, one has as many solutions as in flat space[2]. Because AdS_n is conformally flat the Killing spinors in fact satisfy the conformally invariant equation

$$\nabla_\alpha\gamma_\beta\epsilon + \nabla_\beta\gamma_\alpha\epsilon = \frac{1}{2n}g_{\alpha\beta}\nabla^\sigma\gamma_\sigma\epsilon. \tag{3}$$

which forms much of the basis of “Twistor theory”. Conformal Killing spinors of course arise naturally in conformal supergravity [16]. As a further illustration of historical antecedents, it is interesting to recall that the existence of solutions to an equation of the form (2) was the basic assumption behind the theory of “Wave Geometry” which was extensively developed in Hiroshima in the thirties. The introduction to [38] describing the history of these ideas and the fate of those working on them seems to me to be one the most poignant in the physics literature.

2.4 Three-Branes and Cosmology

In what follows we shall mainly be interested in three-branes. This is partly because they connect with results in four-dimensional quantum field theory. However there is an old tradition of speculation which considers our universe as a three-brane moving in some higher dimensional spacetime (see for example [49]) . Recently this idea has been revived . Cosmologists reading this are cautioned therefore against gratuitously assuming that p-branes have no relevance for their real world.

---

[2] Using the isometric embedding of AdS as an affine quadric that we shall be describing in detail later, the solutions are easilly exhibited as the restriction to the quadric of constant spinors in the flat embedding spacetime.

3 The D-Three-Brane

Now if N gets large the supergravity approximation should get better and better. Consider the case of N three-branes, with N large. This has a supergravity description as a classical BPS spacetime solution of the ten-dimensional Type IIB supergravity theory admitting 16, i.e. half the maximum, Majorana-Weyl, that is real, Killing spinors ϵ [3].

3.1 The Classical Solution

In isotropic coordinates, which are valid only outside the horizon, the solution takes the form

$$ds^2 = H^{-\frac{1}{2}}(-dt^2 + d\mathbf{x}^2) + H^{\frac{1}{2}} d\mathbf{y}^2, \tag{4}$$

where $\mathbf{x} \in \mathbb{E}^3$ is a three vector and $\mathbf{y} \in \mathbb{E}^6$ is a six vector. $H(\mathbf{y})$ is a harmonic function on $\mathbb{E}^6$ and there is also a self-dual five-form

$$F_5 = \star F_5 = dt \wedge dx^1 \wedge dx^2 \wedge dx^3 \wedge d(\frac{1}{H}) + \star(\text{ditto}). \tag{5}$$

The dilaton ϕ is constant

$$e^{2\phi} = g_s. \tag{6}$$

If Yang-Mills fields were present the Yang-Mills coupling constant g_{YM} would be given by

$$g_s = \frac{g^2_{\mathrm{YM}}}{4\pi}. \tag{7}$$

For a solution representing N three-branes located at positions $\mathbf{y}_i$, $i = 1, \ldots, N$, each carrying one unit of five-form magnetic flux one chooses

$$H = 1 + \sum \frac{4\pi g_s \alpha'}{|\mathbf{y} - \mathbf{y}_i|^4}. \tag{8}$$

[3] The reader unfamilar with supersymmetry but willing to accept that eleven-dimensional physics is behind everything may find it helpful to recall that there are *two* inequivalent Clifford algebras Cliff$(10,1)$ each isomorphic to $\mathbb{R}(32)$, the algebra of real 32 by 32 matrices, where one may pick the Clifford representative of the volume form $\gamma_0\gamma_1\ldots\gamma_{10} = \pm 1$. Let us settle on the plus sign. The matrices $\gamma_0, \gamma_1, \ldots, \gamma_9$ generate Cliff$(9,1)$ and one may split the 32 dimensional space S of Majorana spinors into a direct sum $S = S_- \oplus S_+$ of 16 dimensional positive and negative eigenstates of the Clifford representive $\gamma_{10} = \gamma_0\gamma_1\ldots\gamma_9$ of the ten-dimensional volume form. Elements of the summands are called positive or negative chirality Majorana-Weyl spinors. The student with an interest in global matters is invited to reflect on the remarkable effectiveness of this simple piece of mathematics, once one has made the choice of spacetime signature $(10,1)$, and what it implies for spacetimes lacking space or time orientation and what further things it might betoken for mankind. Guidance for the perplexed may be found in [53].

where $\alpha' = l_s^2$ is the Regge slope parameter of string theory and is related to the fundamental string length l_s.

Now let the N branes coalesce. We get

$$H = 1 + (\frac{R}{r})^4, \tag{9}$$

with

$$R = (g_{\mathrm{YM}}^2 N)^{\frac{1}{4}} l_s, \tag{10}$$

and $r = |\mathbf{y}|$.The classical solution is expected to be a good approximation in the limit that N is large but with $\lambda = g_{\mathrm{YM}}^2 N$ held fixed. This corresponds in $U(N)$ gauge theory to a limit the study of which was pioneered by t'Hooft.

3.2 Near Horizon Geometry

Isotropic coordinates break down near the horizon at $r = 0$. For small r the metric tends to

$$(\frac{r}{R})^2(-dt^2 + d\mathbf{x}^2) + \frac{R^2 dr^2}{r^2} + R^2 d\Omega_5^2, \tag{11}$$

where $d\Omega_5^2$ is the standard round metric on S^5 with unit radius.

Now set

$$z = \frac{R}{r} \tag{12}$$

and recall that the standard AdS_{p+2} metric of unit radius in horospheric coordinates $(z, x^\mu)$is given by

$$ds^2 = \frac{1}{z^2}(dz^2 + \eta_{\mu\nu} dx^\mu dx^\nu), \tag{13}$$

with $\mu = 0, 1, \ldots, p$ and $\eta_{\mu\nu}$ is the Minkowski metric. We deduce that the near horizon geometry is that of $AdS_5 \times S^5$ with the two radii of curvature equal. Taking out $1/z^2$ as an overall conformal factor of the limiting ten-dimensional product metric also reveals that it is conformally flat. In fact one may easily extend the argument to show that the metric product of $AdS_r \times S^s$ with radii R_1 and R_2 is conformally flat iff the the two radii of curvature are equal.

Clearly there are considerable advantages associated with horospheric coordinates and we shall be exploiting them further shortly. Before doing so we make a few comments about supersymmetry.

3.3 Supersymmetries

Because it admits a Killing spinor the solution also admits an *everywhere causal* Killing vector field $K^\mu = \bar{\epsilon}\gamma^\mu\epsilon$. In fact the solution has the symmetries expected of a three-brane. The isometry group is $E(3,1) \times SO(6)$ with orbits $\mathbb{E}^{3,1} \times S^5$ [4]. In particular it is locally static, but has degenerate Killing horizons. Near infinity the solution tends to flat ten-dimensional Minkowski spacetime $\mathbb{E}^{9,1}$ which clearly admits the maximum possible, i.e. 32 Majorana-Weyl Killing spinors. Near the horizon the spatial sections have an infinitely long throat resembling that of the familiar extreme Reissner-Nordstrom solution. The solution tends, as we have seen, to the product metric on $AdS_5 \times S^5$, with the two radii of curvature having equal magnitude. This solution also admits 32 Majorana-Weyl Killing spinors and is thus a maximally supersymmetic ground state of Type IIB supergravity theory. In fact it is the basis of a "spontaneous compactification" in which one obtains an effective five-dimensional supersymmetric maximally superysmmetric ground state which is geometrically given by AdS_5. Fluctuations around this solution are given, at the supergravity level, by a five-dimensional gauged supergravity model with gauge group $SO(6)$. Such theories and the properties of such vacua were intensively studied in the past, using just the extensions of Poincaré covariant quantum field theory to the Anti-de-Sitter setting I alluded to above. In the past, the case of AdS_4, usually times S^7 or some other compact seven-dimensional Einstein manifold with positive scalar curvature was of greatest physical interest. However, the lessons learnt then readily generalise.

Remarkably, however, quite unlike the extreme Reissner-Nordstrom solution, the three-brane solution is geodesically complete and everywhere non-singular [33].

3.4 Vacuum Interpolation, Conformal Flatness and Couch-Torrence Symmetry

This phenomenon is referred to as *Vacuum Interpolation* [17]. It is a feature of many other examples. For example the M2-brane of eleven dimensions spatially interpolates between $\mathbb{E}^{10,1}$ and AdS_4 times S^7 and the M5-brane of eleven dimensions spatially interpolates between $\mathbb{E}^{10,1}$ and AdS_7 times S^4. They both admit 16 Killing spinors, but only the latter is everywhere singularity free. The former has singularities very similar to those of Extreme Reissner-Nordstrom. However, neither has another very striking feature of the D3-brane, which it shares with the extreme Reissner Nordstrom solution (RN), in that, because in that case the radii of curvature of the two factors

[4] During the sixties there was an intensive, purely group theoretic, discusssion of the possibility of combining spacetime, $E(3,1)$, and internal Lie group symmetries in some unifying noncompact group G [6]. The upshot was various No-Go Theorems such as those of McGlinn, O'Raifeartaigh and Coleman and Mandula telling one essentially only to consider the direct product of the Poincaré group and a compact semisimple group. This is of course typically what results from Kaluza-Klein theory and other dimensional reduction schemes. It might be interesting to revisit those old ideas in the M-Theory context to see if anything more can be said, given some that the group G must act on a higher dimensional spacetime as an isometry group.

are equal, the metric is conformally flat and has vanishing Weyl tensor. For the M2 and M5 brane, the radii are different and this is not so.

In fact both the D3 and the RN admit an involution which acts by conformal isometries and interchanges the horizon and infinity. For the three-brane the involution is given by

$$r \to \frac{R^2}{r} \tag{14}$$

under which

$$ds^2 \to \left(\frac{R}{r}\right)^2 ds^2. \tag{15}$$

I first became aware of this symmetry from a paper of Couch and Torrence in the Reissner-Nordstrom case [4]. Hence the name I have given is its natural generalisation. In *Schwarzschild* coordinates r in an RN solution of mass M the involution is given by

$$r - M \to \frac{M^2}{r - M}. \tag{16}$$

Of course the isotropic coordinate $|\mathbf{y}| = r - M$ in this case.

It remains unclear whether this symmetry will turn out to play a bigger role in the theory. In other words how, if at all, does this symmetry manifest itself in the quantum theory?

4 AdS_{p+2} and Its Horospheres

The standard definition of AdS_{p+2} is as the quadric M in $\mathbb{E}^{p+1,2}$ with its induced Lorentzian metric given by

$$(X^0)^2 + (X^{p+2})^2 - (X^1)^2 - (X^2)^2 - \ldots - (X^{p+1})^2 = 1. \tag{17}$$

Topologically $AdS_{p+2} \equiv \mathbb{R}^{p+1} \times S^1$, and the isometry group is $O(p+1, 2)$. Later we shall describe the universal covering spacetime $\widetilde{AdS}_{p+1}$.

We remark here that AdS_{p+2} has a natural complexification $M_{\mathbb{C}} \equiv SO(p+3; \mathbb{C})/SO(p+2; \mathbb{C})$ as a complex affine quadric

$$(A + iB)^2 = 1, \tag{18}$$

with $A + iB \in \mathbb{C}^{p+3} = \mathbb{R}^{p+3} + i\mathbb{R}^{p+3}$ in which AdS_{p+2} sits as a real section with $B^1 = B^{p+3} = A^1 = \ldots A^{p+1} = 0$ and $A^1 = X^0$, $A^{p+3} = X^{p+2}$, $B^1 = X^1, \ldots, B^{p+2} = X^{p+1}$. Of course the complexification contains other real sections. What is usually called the "Euclidean section of AdS_{p+2}" is another real section of $M_{\mathbb{C}}$ for which X^0 is pure imaginary and the remaining coordinates are real . This gives hyperbolic space H^{p+2}. For more details about complexified spacetimes and real slices the reader is referred to [1].

Considered as a real $(2p+4)$-dimensional manifold $M_{\mathbb{C}} \equiv TS^{p+2}$, the tangent bundle of the $(p+2)$-sphere. This will be explained in detail later.

To return to AdS_{p+2}, the $\mathbb{Z}_2$ centre of the isometry group is generated from the antipodal map. This is the involution

$$J : X \to -X. \tag{19}$$

By definition $J^2 = \mathrm{id}$. Even though it admits CTCs and indeed closed timelike geodesics (CTGs), nevertheless AdS_{p+2} is time orientable (by deeming that anticlockwise motion in $X^0 - X^{p+3}$ is towards the future for example) and the involution J preserves the time orientation. Anti-de-Sitter spacetime is also space-orientable. If p is even, then J does not preserve space orientation, but, if p is odd, then it does. Now, if $p > 1$, then $O(p+1, 2)$ has four connected components. If p is odd, then the centre J lies in the component connected to the identity. If p is even, then it does not. Thus in the odd case, unless one has good reason, one might expect J to be a gauge symmetry of the theory and one might expect to be able to or to be forced to quotient by J. This is sometimes referred to as the *Elliptic Interpretation*. It would amount to spacetime being the quotient AdS_{p+2}/J. If p is even, then the quotient will not be space orientable. If p is odd, then it will[5]. In any event the way that the quantum representative of J

$$\hat{J} : \mathcal{H}_{\mathrm{qm}} \to \mathcal{H}_{\mathrm{qm}} \tag{20}$$

acts on the quantum mechanical Hilbert space $\mathcal{H}_{\mathrm{qm}}$ is clearly of considerable interest.

Note that exactly parallel remarks apply to the so-called "R-symmetry" group $O(6)$. Total inversion lies in the identity component $SO(6)$ and taking the quotient gives the *orientable* five-manifold $\mathbb{RP}^5 = S^5/\pm 1$.

Horospheric coordinates (z, x^μ) are defined by

$$X^0 + X^{p+1} = \frac{1}{z}, \tag{21}$$

and

$$X^\mu = \frac{x^\mu}{z}, \tag{22}$$

with $\mu = 0, 1, \ldots, p$.

The horospheres are given by $z =$ constant. Each one has the intrinsic geometry of $p+1$ dimensional Minkowski spacetime, just like a flat p-brane. In fact we have a a foliation of AdS_{p+2} by "test" p-branes each one of which is the intersection of the quadric with a null hyperplane in $\mathbb{E}^{p+1,2}$. By $O(p+1, 2)$ symmetry is is easy to see that each horosphere is totally umbilic. In fact, if $p = 3$, one may check that each horosphere solves the equation of motion

[5] In the case of dS_{p+2} the analogue J always reverses time orientation. Passage to the quotient is then disastrous because one is forced to real quantum mechanics

for a test or “probe” D3-brane in this supergravity background, including so-called “Wess-Zumino” terms. Moreover the same is true for the the $r = \text{const}$ surfaces in the exact D3-brane metric.

This gives a rather graphic illustration of how one may think of the solutions as being the result of the superposition of a very large number of light three-branes.

Since

$$J : (z, x^\mu) \to (-z, x^\mu), \tag{23}$$

we need both positive and negative z patches to cover all of AdS_{p+2}. The patches are separated by a Killing horizon at $z = \infty$ which gives rise to a coordinate singularity which is simply the intersection of the quadric with a null hyperplane passing through the origin. Later we will provide a more group theoretic description of horospheres.

4.1 Extension of the Full Three-Brane Metric

This is most simply done [33] by defining

$$z^4 = H = 1 + \left(\frac{R}{r}\right)^4 . \tag{24}$$

Thus

$$\frac{r}{R} = (z^4 - 1)^{-\frac{1}{4}}. \tag{25}$$

The metric becomes

$$ds^2 = \frac{R^2}{z^2}(-dt^2 + d\mathbf{x}^2) + \frac{R^2(dz)^2 z^6}{(z^4-1)^{\frac{10}{4}}} + \frac{R^2 z^2}{(z^4-1)^{\frac{1}{2}}} d\Omega_5^2. \tag{26}$$

This is clearly even in z and the horizon is at $z = -\infty$, but now spatial infinity corresponds to $z = \pm 1$. Using the embedding formula, one may push the exact three-brane metric onto the Anti-de-Sitter metric to give an embedding of the three-brane metric as the proper-subset of $AdS_5 \times S^5$ given by $z^2 > 1$. One may check that $z = 1$ corresponds to a conformal boundary with two connected components analogous to the “Scri” of an asymptotically flat black hole. The entire setup is invariant under the action of the antipodal map J. One may therefore, if one chooses, quotient by J to get a three-brane whose outside and inside are the same!

5 Covering Spaces, the Eternal Return and Wrapping in Time

Many physicists are unhappy with the CTCs in AdS_{p+2} and seek to assuage their feelings of guilt by claiming to pass to the universal covering spacetime

$\tilde{AdS}_{p+2}$. In this way they feel that they have exorcised the demon of "acausality". However therapeutic uttering these words may be, nothing is actually gained in this way. Consider for example the behaviour of test particles. Every timelike geodesic on AdS_{p+2} is a closed curve of the same durations equal to $2\pi R$, which Heraclitus would have called the "Great Year".

In fact all geodesics which depart from a particular event meet up again at the same event after six Great Months. To see this we write the metric in Friedman-Lemaitre-Robertson-Walker form. Geometricaly speaking this is a geodesic normal coordinate system. If $X^0 = \sin t$ and $X^A = T^A \cos t$, where $T^0 = 0$ is a timelike unit vector, $T^A \eta_{AB} T^B = -1$, the metric is

$$ds^2 = -dt^2 + \sin^2 t d\Omega^2_{p+1,-1}, \tag{27}$$

where $d\Omega^2_{p+1,-1}$ is the standard metric on p-dimensional hyperbolic space H^p. Each point on on H^p corresponds to a timelike geodesic. They all start from one event at $t = 0$, reconverge again at $t = \pi$, pass through each other and meet up again in at $t = 2\pi$ and then continue to repeat this cycle for ever. Of course, the metric breaks down at the events $t = \ldots - 2\pi$, $-\pi$, 0,π, 2π,..., but that is because geodesic normal coordinates become singular.

Clearly any observable calculated using timelike geodesics will similarly recur after one Great Year. As far as they are concerned, we are effectively on the identified space. Of course we should look carefully at fluctuations about the background and the boundary conditions to see whether we can have any behaviour which does *not* recur after one great year. We will turn to this point in detail later.

In the meantime we note that, if we pass to the universal covering space $\tilde{D}3$, we may lift the antipodal map and call it $\tilde{J}$. Now $\tilde{J}$ generates an action of the integers taking one asymptotically flat region to infinitely many more. We could, if we wished, identify after any number k of actions of $\tilde{J}$. We shall call this spacetime $D3_k$ and we call the act of identification "wrapping in time".

One situation in which wrapping in time may be advantageous is if we want to identify the spatial coordinates of the three-brane, as would be natural if it were wrapped over a nontrivial cycle in a topologically nontrivial spacetime with a torus factor. The problem is that spatial translations do not act freely. They have fixed points on the horizon. These fixed points would give rise to orbifold singularities if one identified under their action. Because $\tilde{J}$ acts freely, these singularities are eliminated if one composes with some power of $\tilde{J}$, in other words as long as one wraps in time as well as in space.

It is important to distingush between this type of wrapping in time and that obtained by considering the world volume of the three-brane as a so-called " discrete spacetime" of the type considered in the elegant construction of Schild [35,37]. In our terms he considers $\Sigma_4 = \mathbb{E}^{3,1}/L$ where L is the unique Lorentzian self-dual lattice in four dimensions. That model has many attractive features, including invariance under the cover of the discrete Lorentz group $SL(2, \mathbb{G})$, where $\mathbb{G}$ are the Gaussian integers, but would, as should be obvious from the discussion above, lead to orbifold singularities.

6 AdS_{p+2} as a Solvable Group Manifold

It is clear from horospheric coordinates that the Poincaré group $E(p, 1)$ acts on AdS_{p+2}, but obviously not transitively. The largest orbits are the horospheres which are the orbits of the $\mathbb{R}^{p+1}$ group of translations. To get a $(p+1)$-dimensional orbit, one must add the $\mathbb{R}_+$ action referred to for good reaons as the dilatations:

$$x^\mu \to \lambda x^\mu, \tag{28}$$

$$z \to \lambda z, \tag{29}$$

with $\lambda \in \mathbb{R}_+$. The dilatations act on the horospheres. In the embedding space they consist of boosts in the $X^0 - X^{p+1}$ two-plane which take the family of parallel null hyperplanes planes into themselves, but leave invariant the hyperplane passing through the origin which corresponds to the Killing horizon $z \to \infty$.

Clearly the $p+2$ dimensional semidirect product $G_{p+2} = \mathbb{R}_+ \ltimes \mathbb{R}^{p+1}$ acts simply transitively on *one half* of AdS_{p+2} [15]. A convenient matrix representation for $g \in G_{p+2}$ is given by thinking of x^μ as a row matrix and mapping

$$g \to \begin{pmatrix} z & x^\mu \\ 0 & \delta^\mu_\nu \end{pmatrix}. \tag{30}$$

From this a set of left-invariant Cartan-Maurer one forms is easily seen to be given by

$$g^{-1}dg = \begin{pmatrix} z^{-1}dz & z^{-1}dx^\mu \\ 0 & 0 \end{pmatrix}. \tag{31}$$

The AdS_{p+2} metric is clearly left-invariant. Note that, since G_{p+2} is not semisimple, the Killing form of G_{p+1} is singular and does not provide a metric.

Note that G_{p+2} is a subgroup of the causality group $\mathbb{R}_+ \ltimes E(p, 1)$ which, by the Alexandrov-Zeeman theorem [7] [8], is the largest group leaving invariant the causal structure of Minkowski-spacetime $\mathbb{E}^{p,1}$. It is contained in the conformal group $\mathrm{Conf}(p, 1) \equiv O(p+1, 2)/J$ of conformally compactified Minkowski spacetime, but contains only those elements of the latter which leave its conformal boundary “Scri”, $\mathcal{I}$, setwise invariant.

One could systematically develop the theory of AdS_{p+2} using the left-invariant metric on it G_{p+2}, but it seems that this would only give the “outside story” since the orbit of G_{p+2} in AdS_{p+1} contains less than half the space. One can never reach the horizon by acting with the group. Moreover despite the homogeneity of the metric, the group G_{p+2} is geodesically incomplete with respect to the left-invariant metric. In-falling timelike geodesics will penetrate the horizon in finite proper time[6].

This behaviour is rather reminiscent of ancient discussions of the Edge of the Universe Problem and the No-Boundary Proposal by such cosmologists as Archytas[7] and later Nicholas of Cusa. They argued that the universe cannot have a boundary since, if it did, one could always throw a spear towards it. If it had a boundary, then the spear must penetrate, leading to a contradiction. The present example seems to indicate some shortcomings in their logic since, consistent with the homogeneity, the edge of the universe is not actually located at a particular position in G_{p+2}. Nevertheless the spear reaches it in finite proper time.

The moral for us today would seem to be that it is more reasonble to adopt a formalism which covers the horizon. Note that restriction to an orbit of G_{p+2} is definitely not the same as adopting the Elliptic Interpretation. AdS_{p+2}/J, unlike G_{p+2}, *is* geodesically complete. I have never really understood what the slogan "Black Hole Complementarity" means, but possibly this behaviour is an an illustration of what is intended.

The corresponding phenomenon in the case of de-Sitter spacetime is of course the well-known geodesic incompleteness to the past of the Steady State Universe of Bondi, Lyttleton and Hoyle. This may also be thought of as the group manifold G_{p+1}. The many attractive features of this model, its ability to resolve age old philosophical puzzles [25] are due precisely to the group property. The same properties also lead to the physical shortcomings of the model.

6.1 The Iwasawa Decomposition

We are now in a position to view the horospheres in a more abstract light. Consider, to begin with, a noncompact *Riemmanian* symmetric space $X = G/H$, where H is the maximal compact subgroup of the simple but noncompact group G. Then Iwasawa tells us that any element $g \in G$ may be written *uniquely* as

$$g = han, \tag{32}$$

where $h \in H$, $a \in A$ and $n \in N$ where A is abelian and N is nilpotent. The semi-direct product $B = A \ltimes N$ is called the Borel subgroup, that is, one may regard the symmetric space X as the group manifold of B equipped with a left-invariant metric. The orbits of the nilpotent group N are called horospheres. They are labelled uniquely by elements of A and are permuted by elements of H.

[6] This is yet another difference that Lorentzian metrics on noncompact group manifolds can bring about compared with Riemannian metrics

[7] I am grateful to John Barrow for the reference to this Pythagorean from the fifth century BC.

The basic example is n-dimensional hyperbolic space $H^n \equiv SO(n,1)/SO(n)$ which may be regarded as a Wick rotation of AdS_n by taking X^0 to be pure imaginary rather than real. The horospheric coordinate t is then pure imaginary. This is the upper half space model of hyperbolic space since $z > 0$. One has $G = SO(n,1)$, $H = SO(n)$, $A = \mathbb{R}_+$, the dilatations, and $N = \mathbb{R}^{p+1}$, the translations. The Iwasawa coordinates are global: they cover all of hyperbolic space.

As we have seen, the case of $AdS_n = SO(n,2)/SO(n.1)$ is similar, except that the Iwasawa coordinates are not global: they do not cover all of AdS_n.

6.2 Symmetric Space Duality, the Anti-Hopf Fibration and the Goedel Viewpoint

The horosphere concept has a another interesting application to the geometry for AdS_n in the case that $n, = 2m+1$, is odd. It is illuminating to place the construction in a general context, so we begin by recalling that to every noncompact Riemannian symmetric space $X = G/H$ there is associated a compact symmetric space $\hat{X} = \hat{G}/H$. If the Lie algebra of G is $\mathrm{g} = \mathrm{h} \oplus \mathrm{p}$ then the Lie algebra of $\hat{G}$ is $\hat{\mathrm{g}} = \mathrm{h} \oplus i\mathrm{p}$. Thus the noncompact generators p of the noncompact group G have become the compact generators ip of the compact group $\hat{G}$. The Riemannian symmetric space X is topologically trivial and carries an Einstein metric with negative scalar curvature. The dual Riemannnian symmetric space is topologically nontrivial and carries an Einstein metric with positive scalar curvature. For example $\hat{H^n} = S^n$. We can obviously define the inverse map so that for example $\hat{S^n} = H^n$.

Now choose $\hat{X} = SU(m+1)/U(m) \equiv \mathbb{CP}^m$ which is the base manifold of the Hopf fibration of S^{2m+1} by S^1,

$$\mathbb{CP}^m = S^{2m+1}/U(1). \tag{33}$$

Explicitly, $S^{2m+1} \subset \mathbb{C}^{m+1} \equiv \mathbb{E}^{2m+2}$ is given by

$$|Z^1|^2 + \ldots + |Z^{m+1}|^2 = 1, \tag{34}$$

where Z^a, $a = 1, \ldots m+1$ are complex affine coordinates for $\mathbb{C}^{m+1} \equiv \mathbb{E}^{2m+2}$. The $U(1)$ action is

$$Z^a \to e^{i\alpha} Z^a. \tag{35}$$

Now let us pass to the symmetric space dual of this construction. We replace S^{2m+1} by $AdS_{2m+1} \subset \mathbb{C}^{m+1} \equiv \mathbb{E}^{2m,2}$ which is given by

$$-|Z^1|^2 - \ldots + |Z^{m+1}|^2 = 1. \tag{36}$$

Thus the $U(1)$ action is as before, but now it has *timelike* circular orbits in AdS_{2m+1}, i.e. the orbits are CTCs and therefore the base space has a Riemannian metric. In fact $X = SU(m,1)/U(m) \equiv H^m_{\mathbb{C}}$ is the unit ball in $\mathbb{C}^m$

equippped with the Bergman metric, which is the dual of the Fubini-Study metric on $\mathbb{CP}^m$. Both are homogeneous Einstein-Kähler 4-metrics, and as such examples of Gravitational Instantons. One has positive cosmological constant and the other has negative cosmological constant. In fact the Bergman metric is the infinite NUT charge limit of the Taub-NUT-Anti-de-Sitter metrics [12].

The metric is

$$ds^2 = -(dt + A_i dx^i)^2 + g_{ij} dx^i dx^j, \tag{37}$$

where $i = 1, 2, \ldots, 2m$, g_{ij} is the Einstein-Kähler metric and dA is the Kähler form.

In traditional relativist's language, AdS_{2m+1} has been exhibited a stationary metric with constant Newtonian potential $U = \frac{1}{2}\log(-g_{00})$. The Coriolis or gravitomagnetic connection, governing frame-dragging effects, corresponds precisely to the connection of the standard circle bundle over the Kähler base space. The curvature is the Kähler form. In fact one may replace the Bergman manifold with any other $2m$ dimensional Einstein-Kähler manifold with negative scalar curvature and obtain a $(2m+1)$-dimensional Lorentzian Einstein manifolds admitting Killing spinors in this way.

The general metric is

$$ds^2 = -(dt + A_i dx^i)^2 + g_{ij} dx^i dx^j, \tag{38}$$

where $i = 1, 2, \ldots, 2m$, g_{ij} is the Einstein-Kähler metric and dA is the Kähler form. The timelike coordinate t is periodic with period 2π. It would seem that there should be applications here to the study of rotation and the AdS/CFT correpondence [30]. A point of interest is that Fourier analyzing the mode QFT mode functions on the spacetime gives rise a to Geometric Quantization problem on the Kähler base manifold. A related construction, not using a Kähler base, providing higher dimensional analogues of the Lorentzian Taub-NUT metric is given in [11].

The simplest case is $m = 1$ which is closely related to the Goedel Universe. In this case the base space is the two-dimensional real hyperbolic space H^2 and the Bergman metric is the standard Poincaré metric.

Geometricaly the Goedel universe has a product metric on $\mathbb{R} \times SL(\tilde{2}, \mathbb{R})$. For our purposes it is more convenient to pass down to $SL(2, \mathbb{R})$. Now equipped with its biinvariant or Killing metric one has:

$$SL(2, \mathbb{R}) \equiv AdS_3 \tag{39}$$

and

$$AdS_3/J = SO(2, 1). \tag{40}$$

In terms of a left invariant basis the biinvariant metric

$$ds^2 = \frac{1}{4}(\sigma_1^2 + \sigma_2^2 - \sigma_0^2). \tag{41}$$

The anti-Hopf fibres have a time like tangent vector dual to the one-form σ_0.

Goedel himself did not choose the biinvariant metric but rather a left invariant metric on $\tilde{SL(2,\mathbb{R})}$ which is "locally rotationally symmetric", that is, invariant under the right action of $U(1)$. This right action commutes with a left action of a circle subgroup of $SL(2,\mathbb{R})$. His metric is

$$ds^2 = \frac{1}{4}(\sigma_1^2 + \sigma_2^2 - \lambda^2 \sigma_0^2), \tag{42}$$

where λ is a constant appropriately chosen to solve the Einstein field equations for rigidly rotating dust. Note that $\sigma_1^2 + \sigma_2^2$ is the standard metric on H^2.

6.3 Heisenberg Horospheres, Finite in All Directions

If we think of $H_{\mathbb{C}}^m$ as the noncompact symmetric space $SU(m,1)/U(m)$, it also admits a horospherical or Iwasawa decompostion. The abelian factor A is again $\mathbb{R}_+$. The nilpotent factor N is now a Heisenberg group [12]. Thus for example, in addition to the standard foliation, $AdS_5 \equiv U(2,1)/U(2)$ also admits a foliation by a one parameter family consisting of the time-like world volumes of 3-branes. Now, because t is periodic, these rotating 3-branes have a periodic time coordinate. They are "wrapped in time".

What about "wrapping in space"? A related question is whether there is a freely acting discrete subroup $\Gamma \subset SO(n-1,2)$ acting properly discontinuously on AdS_n such that AdS_n/Γ is compact. For reasons connected with the Lorentzian Gauss-Bonnet Theorem, this is only possible if $n = 2m+1$ is odd. In that case there are many suitable $2m+1$ dimensional lattices $L \subset U(m,1)$ [20]. Thus indeed one may wrap branes in both space and time in AdS_5. Moreover, because of the holomorphic nature of the construction, the wrapping should be compatible with superysmmetry.

The resultant nonsingular compact Lorentzian spacetimes have no boundary and will certainly have CTCs, but may well prove interesting in the context of string theory where compact flat spacetimes have already been analysed [19]. moreover partially compactified AdS models have already been used to investiagtae cosmological aspects of the AdS/CFT correspondence [23].

Interestingly, it is an old result of Calabi and Markus that there are *no* compact quotients of de-Sitter spacetimes without boundary in any dimension. The best one may do is to identify by the antipodal map to get a de-Sitter spacetime with one, rather than the usual past and future boundaries. However, as mentioned earlier, this destroys the time orientation and seems to be fatal quantum mechanically [31].

6.4 Horospheric Brane-Waves

There is an analogue of the pp-wave metrics which represents gravitational waves propagating in Anti-de-Sitter spacetime which I worked out with Stephen Siklos several years ago (see [43] for details and references). The metrics are conformal to pp-waves. They may be used to construct p-branes on which propagate gravitational waves. Actually the following $(p+2)$ dimensional metric is slightly more general

$$ds^2 = \frac{1}{z^2}\left\{-dudv + H(u,z,x^a)du^2 + dz^2 + g_{ab}(x^a)dx^a dx^b\right\}. \tag{43}$$

This will satisfy the Einstein equations with cosmological constant as long as

$$R_{ab} = 0 \tag{44}$$

and

$$z^p \left(\frac{\partial}{\partial z}\left(\frac{1}{z^p}\frac{\partial H}{\partial z}\right)\right) + \nabla_g^2 H = 0, \tag{45}$$

where $a, b = i, 2, \ldots, p-1$ and ∇_g^2 is the Laplacian with respect to the metric g_{ab}. The dependence on u is arbitrary. If the metric g_{ab} is flat, i.e. if $g_{ab} = \delta_{ab}$, then the metric is conformal to a pp-wave. It will then admit half the maximum number of Killing spinors, i.e. those which satisfy

$$\bar\epsilon\gamma^\mu\epsilon\frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial v}. \tag{46}$$

The right hand side of (46) is a lightlike Killing vector field.

7 Conformal Compactifications and the Boundary of AdS_{p+2}

The basic observation behind the AdS/CFT correspondence is the statement that the conformal boundary of AdS_{p+2} is a (twofold cover of) conformally compactified Minkowski spacetime $\overline{\mathbb{E}^{p,1}}$, that is,

$$\partial(AdS_{p+2}) = S^p \times S^1, \tag{47}$$

or lifting to the universal cover

$$\partial(Ad\tilde{S}_{p+2}) \equiv ESU_{p+1}, \tag{48}$$

where $ESU_{p+1} \equiv S^p \times \mathbb{E}^{0,1}$ is the Einstein static universe. Indeed $\tilde{A}dS_{p+2}$ is conformally flat and may be conformally embedded into one half of ESU_{p+2}. It is more or less obvious that the conformal boundary is a copy of ESU_{p+1}.

The main idea of Maldacena is that, since the isometry group of a manifold, referred to in this context as the "bulk", is the conformal isometry group of its conformal boundary, then Conformal Field Theory on the boundary should, in the large N limit, be equivalent to Type IIB string theory in the interior. The idea is obviously capable of further elaborations and generalisations which I shall not enter into here.

We shall start by describing the compactification of Minkowski spacetime and then that of Anti-de-Sitter spacetime.

7.1 Conformally Compactified Minkowski Spacetime

If we adjoin to the causality group of $p+1$ dimensional Minkowski spacetime the special conformal transformations

$$x^\mu \to \frac{x^\mu + c^\mu x^2}{1 + 2c_\mu x^\mu + c^2 x^2} \tag{49}$$

we obtain the full conformal group $\mathrm{Conf}(p,1) \equiv SO(p+1,2)/\mathbb{Z}_2$. This isomorphism is easily verified at the Lie algebra level but globally things are more subtle. The conformal group acts not on Minkowski spacetime but its conformal compactification $\overline{\mathbb{E}^{p,1}} \equiv (S^p \times S^1)/\mathbb{Z}_2$. To see this, we identify $\overline{\mathbb{E}^{p,1}}$ with the space of null rays in $\mathbb{E}^{p+1,2}$. We recover Minkowski spacetime by intersecting with the "light cone" with the null hyperplane

$$X^0 + X^{p+1} = \frac{1}{z}. \tag{50}$$

The stability group of the null hyperplane is just the Poincaré group $E(p,1)$. The null hyperplane captures some but not all of the possible light rays. We miss those parallel to the null hyperplane. These points on the conformal boundary of Minkowski spacetime which is usually called "Scri", standing for script i, $\mathcal{I}$. The entire set of light rays constitutes an $(S^p \times S^1)/\mathbb{Z}_2$.

The usual picture introduced by Penrose is slightly different. It is obtained by regarding the conformal compactification $\{\bar{M}, \bar{g}\}$ of a manifold $\{M, g\}$ as a compact manifold with boundary $\partial\bar{M}$, conformally embedded in some larger manifold $\{\tilde{M}\hat{g}\}$. On $M = \bar{M} \setminus \partial\bar{M} \subset \hat{M}$ one has $\hat{g} = \Omega^2 g$, where Ω is a smooth function on $\tilde{M}$ which vanishes on ∂M but such that $d\Omega \neq 0$ on ∂M. Thus Ω vanishes as the distance from the boundary.

Thus Minkowski spacetime in spherical polars has the metric

$$ds^2 = -du dv + r^2 d\Omega^2_{p-1}, \tag{51}$$

where $u = t - r$ and $v = t + r$ are retarded and advanced null coordinates. If one sets $u = \tan(\frac{T-\chi}{2})$ and $v = \tan \frac{(T+\chi}{2})$, one gets

$$ds^2 = \Omega^{-2}(dT^2 + d\chi^2 + \sin^2 d\Omega^2_{p-1}) \tag{52}$$

with $\Omega = 2\cos(\frac{T-\chi}{2})\cos(\frac{T+\chi}{2})$. One sees that

$$d\Omega^2_p = d\chi^2 + \sin^2 \chi d\Omega^2_{p-1} \tag{53}$$

is the metric on the unit p-sphere with $0 \leq \chi \leq \pi$. Thus the universal cover of the conformal compactification of Minkowski spacetime is the Einstein Static universe $ESU_{p+1} \equiv S^p \times \mathbb{E}^{0,1}$. In fact according to a result of Schmidt [63] ESU_{p+1} is maximal in the sense that it cannot be conformally embedded into a a strictly larger manifold. Thus a an open conformally flat $(p+1)$-dimensional manifold, such as $H^p \times \mathbb{E}^{0,1}$ for example, may typically be conformally embedded into ESU_{p+1} as a (possibly proper) subset. This is a standard construction, due to Penrose, for Friedman-Lemaitre-Robertson-Walker universes. We shall use it later when dealing with black holes with exotic topologies.

The involution $\hat{J}$ acts as

$$\hat{J} : (T, \chi, \mathbf{n}) \to (T + \pi, \pi - \chi, -\mathbf{n}). \tag{54}$$

Thus it consists of a time shift by six Great Months, i.e. half a Great Year, composed with the antipodal map on the S^p factor. It therefore identifies what is usually called $\mathcal{I}^+ \equiv v = \infty \equiv T + \chi = \pi$ with $\mathcal{I}^- \equiv u = -\infty \equiv T - \chi = -\pi$. A light ray passing through $\mathcal{I}^+$ should thus reappear passing through $\mathcal{I}^-$.

Of course in the context of conventional macroscopic physics this is ridiculous and clearly does not happen. However, there may well be circumstances when considering the AdS/CFT correspondence, for example, in which the compactified boundary conditions are appropriate.

Consider for example an experimental colleague in the laboratory investigating the steady state configuration of a physical system which is being periodically excited, such as a resonance. The correct boundary conditions for a theorist to use to describe the resonating system are those of the Eternal Return with Great Year equal to to the inverse frequency of the resonance. There is, in that case, no question that time "really is" periodic.

In the special case of four-dimensional Minkowski spacetime there is an alternative and some times more useful description (see e.g. [39] for details and references) which starts with thinking of the points x of Minkowski spacetime as two by two Hermitian matrices, i.e $x \in u(2)$ the Lie algebra of $U(2)$. The compactification corresponds to passing to the group by means of the Cayley map

$$x \to U = (1 + ix)(1 - ix)^{-1}. \tag{55}$$

Thus $\overline{E^{3,1}} \equiv U(2)$. The metric, which is just the obvious invariant metric $-\mathrm{Tr} U^{-1} dU U^{-1} dU$ is of course *Lorentzian* . The $U(1)$ factor is timelike. Thus the two fold cover is $SU(2) \times U(1)$ and the universal cover is $SU(2) \times \mathbb{R}$. A similar construction will work for the reals and the quarternions in two and six spacetime dimensions.

In other dimensions there is a related construction using Clifford algebras $x = \gamma_\mu x^\mu$.

7.2 The Conformal Compactification of AdS_{p+3}

The embedding of AdS_{p+1} is given by

$$X^0 = \sqrt{1+r^2}\sin t \tag{56}$$

$$X^{p+3} = \sqrt{1+r^2}\cos t, \tag{57}$$

$$X^i = r\sin\chi n^i, \tag{58}$$

$$X^{p+1} = r\cos\chi. \tag{59}$$

This also gives a conformal embedding into ESU_{p+2} because the metric is

$$\Omega^{-2}\{dt^2 + d\omega^2 + \sin^2\omega(d\chi^2 + \sin^2\omega d\Omega^2_{p-1})\} \tag{60}$$

, where $\Omega^2 = \cos\omega$ and $r = \tan\omega$. Since spatial infinity, $r = \infty$, corresponds to $\omega = \frac{\pi}{2}$ the conformal boundary of $\tilde{A}dS_{p+2}$ is the timelike cylinder ESU_{p+1} as advertised. To get AdS_{p+2} we must identify t modulo 2π. From (54), it is clear that its boundary is the twofold cover of the set of null rays, i.e. of $\overline{\mathbb{E}^{p,1}}$. The latter is the boundary of AdS_{p+2}/J.

Note that, if one adopts horospheric coordinates, one might have concluded that the conformal boundary of AdS_{p+2} is a copy of Minkowski spacetime $\mathbb{E}^{p,1}$ situated at $z = 0_+$. However, this is clearly only part of the boundary. Recalling that the other side of the horizon has z negative, one might then try to add in another copy situated at $x = 0_-$. However, this leads to overcounting. One must identify points related by inversions

$$x^\mu \to \frac{x^\mu}{x^2}. \tag{61}$$

Roughly speaking, one has to attach to Minkowski spacetime the lightcone of the origin. This corresponds to $\mathcal{I}$. However, care must be taken with signs and the upshot is that one lands up on $S^p \times S^p/\mathbb{Z}_2$ again.

7.3 The Conformal Boundary of H^{p+1} and the Doppelganger on the Other Sheet

Superficially, using the horospheric, or upper half space, representation of the metric

$$ds^2 = \frac{1}{z^2}(dz^2 + d\mathbf{x}^2_{p+1}), \tag{62}$$

one might have concluded that the conformal boundary of hyperbolic space is $\mathbb{E}^{p+1}$ situated at $z = 0$, but this leaves out a single point at $z = \infty$. The boundary is actually S^{p+1}. This is most simply seen by thinking of the H^{p+2} as the set of future directed timelike lines passing through the origin of $\mathbb{E}^{p+1,1}$. If one cuts this with a spacelike hyperplane at unit distance, the rays are captured inside a ball of unit radius. The bounding $p+1$ sphere corresponds to the null rays through the origin. The detailed calculation is very similar to the standard case of stereographic projection. In spherical coordinates the hyperbolic metric is

$$ds^2 = d\omega^2 + \sinh^2 \omega d\Omega^2_{p+1}. \tag{63}$$

If $r = \tanh(\frac{\omega}{2})$, this becomes

$$ds^2 = \frac{4}{(1-r^2)^2}(dr^2 + r^2 d\Omega^2_{p+1}). \tag{64}$$

One therefore has $\Omega = \frac{1}{2}(1-r^2)$ which vanishes as the distance on the boundary $r = 1$.

There is an analogue of the antipodal map for hyperbolic space, reflection in the origin of Minkowski spacetime. However, it takes one from the upper sheet of future directed timelike lines to the disconnected lower sheet of past directed timelike lines. One might have thought therefore that the involution plays no role in the "physical sheet". However, this is *not* so. When constructing "Euclidean" Green's functions inside the unit ball one must choose between Dirichlet or Neumann boundary conditions. Calculation reveals that in order to incorporate this it is necessary to add an image source to the direct contribution coming from a Doppelganger on the other sheet and whose strength is equal in magnitude to that of the direct source and whose sign determines whether one has Dirichlet or Neumann case.

To see this explicitly we first introduce the chordal distance $\sigma(x, x')$ of two points on AdS_{p+2} or its complexification

$$X^A \eta_{AB} X^B = -1. \tag{65}$$

In terms of the embedding coordinates one has:

$$\sigma = -\frac{1}{2}\eta_{AB}(X^A - X'^A)(X^B - X'^B). \tag{66}$$

It follows that

$$\sigma = 1 + X^A \eta_{AB} X'^B. \tag{67}$$

Obviously $\sigma = 0$ if the points coincide and $\sigma = 2$ if they are antipodal, i.e. $X^A = -X'^A$.

In horospheric coordinates one has

$$\sigma = \frac{(x^\mu - x'^\mu + (z - z')^2)^2}{2zz'}. \tag{68}$$

For a scalar field of mass m one defines

$$a = \frac{p+1}{2} + \sqrt{(\frac{p+1}{2})^2 + m^2}, \tag{69}$$

$$b = \frac{p+1}{2} - \sqrt{(\frac{p+1}{2})^2 + m^2} \tag{70}$$

$$c = \frac{p+1}{2}. \tag{71}$$

The free two-point correlation functions may be expressed in terms of hypergeometric functions and, in the Dirichlet case, are proportional to

$$\sigma^{-a} F(a, a+1-c, a+1-b; \frac{2}{\sigma}). \tag{72}$$

One gets the Neumann case by interchanging the the roles of a and b, i.e. taking the opposite sign for the square root in all formulæ. The square roots remain positive even if m^2 is negative, but not too negative. This is the Breitenlohner-Freedman bound.

The hypergemetric function has poles at zero, one and infinity. The first occurs when the points coincide, the second when they are antipodal. The third when they they have infinite separation.

8 The Geodesic Flow on AdS and the Future Tube of the Boundary

If one is interested in quantizing a relativistic particle moving in AdS_n, one approach is to look at the relativistic phase space $T^\star AdS_n$, pass to the constrained space and then to "quantize"it. Because of the high symmetry, one is able to give a rather explicit description of the relevant spaces in group-theoretic terms. They turn out to have some striking properties.

Recall that, in general, the relativistic phase space of a spactime M is the cotangent bundle $T^\star M$ with coordinates $\{x^\mu, p_\mu\}$, canonical one-form $p_\mu dx^\mu$ and symplectic form

$$\omega = dp_\mu \wedge dx^\mu. \tag{73}$$

The geodesic flow is generated by the covariant Hamiltonian

$$\mathcal{H} = \frac{1}{2} g^{\mu\nu} p_\mu p_\nu. \tag{74}$$

The flow for a timelike geodesic, corresponding to a particle of mass m, lies on the level sets, call them Γ, given by

$$\mathcal{H} = -\frac{1}{2} m^2. \tag{75}$$

Locally at least, one may pass to the reduced phase space $P = \Gamma / G_1$, where G_1 is the one-parameter group generated by the covariant Hamiltonian, $\mathcal{H}$, by a "Marsden-Weinstein reduction". Geometrically, the group G_1 takes points and their cotangent vectors along the world lines of the timelike geodesics.

The reduced $(2n-2)$-dimensional phase space P is naturally a symplectic manifold and one may now attempt to implement the geometric quantization programme by "quantizing " P.

In the general case it seems to be difficult to carry out this procedure and compare it with the results of more conventional quantum field theory approaches because one does not have a good understanding of the space of timelike geodesics P. In the case of AdS_n, however, the space may be described rather explicitly. It turns out to be a Kähler manifold which is isomorphic to the future tube T^+_{n-1} of $(n-1)$-dimensional Minkowski spacetime.

In AdS_n every timelike geodesic is equvalent to every other one under an $SO(n-1,2)$ transformation. They may all be obtained as the intersection of some totally timelike 2-plane passing through the origin of of the embedding space $\mathbb{E}^{n-1,1}$ with the AdS_n quadric. The space P of such two planes may thus be identifed with the space of geodesics. It is a homogeneous space of the isometry group. In fact it is the Grassmannian $SO(n-1,2)/(SO(2)\times SO(n))$. Note that, as one expects, the dimension of P is $2n-2$. The denominator of the coset is the maximal compact subgroup of $SO(n-1,2)$. Two factors correspond to timelike rotations in the timelike 2-plane and rotations of the normal space respectively. The former may be identified with the one parameter group G_1 generated by the covariant Hamiltonian $\mathcal{H}$. Thus the level sets Γ is the coset space $SO(n-1,2)/SO(n)$.

Now the striking fact is that the reduced phase space $P\equiv SO(n-1,2)/(SO(2)\times SO(n))$ coincides with one of the four series of irreducible bounded symmetric domains, first classified by Cartan [57]. Our case is Ω^{IV}_{n-1} which, as mentoned above, may also be identified with the *Future Tube* T^+_{n-1} of $(n-1)$-dimensional Minkwoski spacetime $\mathbb{E}^{n-1,1}$. This space plays a central role in quantum field theory in flat spacetime since Wightman functions and Green's functions are typically boundary values of holomorphic functions on the future tube. The future tube is defined as those complex vectors $z\in\mathbb{C}^{n-1}$ whose imaginary part lies in the future lightcone.

The space P caries a natural Einstein Kähler metric. The complex structure is given by the $SO(2)$ action. One may regard the Kähler form as the curvature of a circle bundle. This bundle is the constraint manifold Γ. Actually the entire cotangent bundle $T^\star AdS_n$, which is a 2-plane bundle over P, carries a Ricci-flat pseudo-Kähler metric. This this metric has signature $(2n-2,2)$. The timelike coordinates correspond to the time around circle direction, and a coordinate labelling the levels sets $2\mathcal{H}=-m^2$.

The existence of this Ricci-flat pseudo-Kähler metric may be obtained by analytically continuing Stenzels's positive definite Ricci-flat Kähler metric on the cotangent bundle of the standard n-sphere, $T^\star S^n$ [59]. The simplest case is when $n=2$. Stenzel's metric is then the Eguchi-Hanson metric which may be analtyically continued to give a "Kleinian" metric of signature $(2,1)$ on $T^\star AdS_2$.

As noted earlier, $T^\star S^n$ may be identified with an affine quadric in $\mathbb{C}^{n+1}$. This may be seen as follows: $T^\star S^n$ consist of a pair of real $(n+1)$ vectors X^A and P^A such that

$$X^1X^1+X^2X^2+\ldots+X^{n+1}X^{n+1}=1, \tag{76}$$

$$X^1P^1 + X^2P^2 + \ldots + X^{n+1}P^{n+1} = 0. \tag{77}$$

If $P = \sqrt{P^1P^1 + P^2P^2 + \ldots + P^{n+1}P^{n+1}}$, one may map $T^\star S^n$ into the affine quadric

$$(Z^1)^2 + (Z^2)^2 + \ldots + (Z^{n+1})^2 = 1 \tag{78}$$

by setting

$$Z^A = A^A + iB^A = \cosh(P)X^A + i\frac{\sinh(P)}{P}P^A. \tag{79}$$

Stenzel then seeks a Kähler potential depending only on the restriction to the quadric (18) of the function

$$\tau = |Z^1|^2 + |Z^2|^2 + \ldots + |Z^{n+1}|^2. \tag{80}$$

The Monge-Ampère equation now reduces to an ordinary differential equation.

In the case of AdS_{p+2} we may proceed as follows. The bundle of future directed timelike vectors in AdS_{p+2}, T^+AdS_{p+2} consists of pairs of timelike vectors X^A, P^A in $\mathbb{E}^{p+1,2}$ such that

$$X^AX^B\eta_{AB} = -1 \tag{81}$$

and

$$X^AP^B\eta_{AB} = 0, \tag{82}$$

with P^A future directed and $\eta_{AB} = \mathrm{diag}(-1, -1, +1, \ldots, +1)$ the metric. We define $P = \sqrt{-P^AP^B\eta_{AB}}$ and

$$Z^A = \cosh(P)X^A + i\frac{\sinh(P)}{P}P^A \tag{83}$$

which maps T^+AdS_{p+2} to the affine quadric

$$Z^AZ^B\eta_{AB} = -1. \tag{84}$$

One then seeks a Kähler potential depending only on the restriction to the quadric (18) of the function

$$\tau = |Z^0|^2 + |Z^{p+2}|^2 - |Z^1|^2 - \ldots - |Z^{p+1}|^2. \tag{85}$$

The Monge-Ampère equation again reduces to an ordinary differential equation.

We return to the reduced phase space P. It may be realised as a bounded domain $D \subset \mathbb{C}^{n-1}$ and as such it has a $(2n-1)$-dimensional topological boundary ∂D. More interestingly, lying inside this topological boundary, ∂D is its $(n-1)$-dimensional *Shilov boundary* S. If $w \in \mathbb{C}^{n-1}$ is a complex $(n-1)$ column vector and $w^2 = w^tw$ and $|w|^2 = w^\dagger w$, then the domain D is defined by [58]

$$1 - |w|^2 \geq \sqrt{|w|^4 - |w^2|^2}. \tag{86}$$

The topological boundary is given by the real equation:

$$1 - 2|w|^2 + |w^2|^2 = 0. \tag{87}$$

On the other hand, the Shilov boundary is determined by the property that the maximum modulus of any holomorphic function on P is attained on S. Consider, for example, the holomorphic function w. It atttains its maximum modulus when $w = \exp(i\theta)\mathbf{n}$, where $\mathbf{n}$ is a real unit $(n-1)$ vector. Thus S is given by $S^1 \times S^{n-1}/\mathbb{Z}_2$.

It is no coincidence that S is topologicaly the same as the conformal boundary of AdS_n. To see why, following Hua, who refers to D as "Lie Sphere Space" we can linearise the action of $SO(n-1,2;\mathbb{R})$ by embedding D into $\mathbb{C}^{n+1}$. Let

$$W^0 - iW^{n+1} = \frac{1}{u}, \tag{88}$$

$$W^0 + iW^{n+1} = \frac{w^2 2}{u} \tag{89}$$

and

$$W^i = \frac{w^i}{u}, \tag{90}$$

where $i = i, \ldots, n-1$ and the complex, horospheric type coordinate u should be set to unity to recover D. The n coordinates (u, w^i) thus parameterise the complex lightcone, i.e. the real $2n$ dimensional submanifold $W \subset \mathbb{C}^{n+1}$ given by

$$(W^0)^2 + (W^{n+1})^2 - W^2 = 0. \tag{91}$$

The domain D consists of rays through the origin lying in W, that is, one must identify rays W^A and λW^A, where $\lambda \in \mathbb{C}^\star \equiv \mathbb{C} \setminus 0$. Thus $D = W/\mathbb{C}^\star$.

Evidently $SO(n-1,2;\mathbb{R})$ acting in the obvious way on $\mathbb{C}^{n+1}$ leaves W invariant and commutes with the $\mathbb{C}^\star$ action. Thus the action of $SO(n-1,2;\mathbb{R})$ descends to D. If we restrict the coordinates W^A to be real, we obtain the standard construction of $(n-1)$-dimensional compactified Minkowski spacetime as light rays through the origin of $\mathbb{E}^{n-2,2}$.

The case $n = 4$ is special since $SO(4,2) \equiv SU(2,2)/\mathbb{Z}_2$. This leads to the equivalence of $\Omega^I_{2,2}$ and Ω^{IV}_4. As mentioned above, one may identify points in real four-dimensional Minkowski spacetime $\mathbb{E}^{3,1}$ with two by two Hermitian matrices $x = x^0 + \mathbf{x} \cdot \sigma$. The future tube T^+_4 then corresponds to complex matrices $x = z^0 + \mathbf{z} \cdot \sigma$ with imaginary part positive definite. The Cayley map

$$z \to w = (z-i)(z+i)^{-1} \tag{92}$$

maps this into the bounded holomorphic domain in $\mathbb{C}^4$ consisting of the space $\Omega^I_{2,2}$ of two by two complex matrices w satisfying

$$1 - ww^\dagger > 0. \tag{93}$$

For more details, the reader is directed to [60]. For this approach to the compactification of Minkowski spacetime see also [61,62].

9 The Anti-de-Sitter Algebra and Quantized Energies

If a Lie group G with structure constants $C_a{}^b{}_c$ acts on the left on a manifold M, the Killing vector fields $\mathbf{K}_a$ have Lie brackets

$$[\mathbf{K}_a, \mathbf{K}_c] = -C_a{}^b{}_c \mathbf{K}_b. \tag{94}$$

In quantum mechanics one often prefers to work with $\hat{M}_a = -i\mathbf{K}_a$ acting on spacetime scalar fields as a formally self-adjoint operator with respect to the inner product obtained by integrating over spacetime. Clearly

$$[\hat{M}_a, \hat{M}_c] = iC_a{}^b{}_c \hat{M}_b. \tag{95}$$

The AdS_{p+2} group $SO(p+1,2)$ corresponds to

$$\mathbf{K}_{AB} = X_A \partial_B - X_B \partial_A \tag{96}$$

and therefore

$$[\hat{M}_{AB}, \hat{M}_{CD}] = i\hat{M}_{AC}\eta_{BD} - i\hat{M}_{BC}\eta_{AD} + i\hat{M}_{AD}\eta_{BC} - i\hat{M}_{BD}\eta_{AC}. \tag{97}$$

Upper case Latin indices run from 0 to $p+2$ and $\eta_{AB} = \mathrm{diag}(-1, +1, \ldots, +1, -1)$. Greek indices run from 0 to p. Lower case Latin indices run from 1 to p.

The maximal compact subgroup of $SO(p+1,2)$ is $SO(p) \times SO(2)$ with generators $\hat{M}_{ij}$ and $\hat{M}_{0,p+2}$. The latter corresponds to rotations in the totally time-like $X^0 X^{p+2}$ plane. The associated Killing vector field is the globally static Killing field such that, in adapted coordinates, the metric is

$$ds^2 = -(1+r^2)dt^2 + \frac{dr^2}{1+r^2} + r^2 d\Omega^2_{p-1} \tag{98}$$

with $0 \le t \le 2\pi$.

In the case of dS_{p+2} and $SO(p+2,1)$, X^{p+2} would be spacelike and the maximal compact subgroup would be $SO(p+2)$. It that case $M_{0,p+2}$ would be a noncompact generator corresponding to a boost. The associated Killing vector is not globally static as is clear form the metric in adapted coordinates:

$$ds^2 = -(1-r^2)dt^2 + \frac{dr^2}{1-r^2} + r^2 d\Omega^2_{p-1} \tag{99}$$

with $-\infty < t < \infty$. There is a Killing horizon at $r = 1$. This difference is crucial for our concept of energy at the classical and the quantum level.

In the de-Sitter case there is no useful global energy concept. As Wigner first realised, there are no "positive energy" represention of $SO(p+2,1)$ [29]. The point is that one may easily find a diagonal element of the identity component of $SO(p+2,1)$, call it g, such that under the adjoint action

$$\hat{M}_{p+2\,0} \to g\hat{M}_{p+2\,0}g^{-1} = -\hat{M}_{p+2\,0}. \tag{100}$$

The existence of g means that in any unitary representation $\hat{U}(g)$ acts on an energy eigenstate $|E\rangle$ with energy E to give a new state $\hat{U}|E\rangle$ with energy $-E$. Acting on de-Sitter spacetime the element g takes one from one side of the event horizon to the other. This observation is closely related to the thermal emission from cosmological event horizons [36] which Hawking and I disovered in complete ignorance of Wigner's prescient observation.

Wigner's observation is also related to the fact that de-Sitter backgrounds break supersymmetry. Being conformally flat they certainly admit a full set of solutions ϵ of the twistor equation (3). However, the causal vector fields $\bar{\epsilon}\gamma^{\mu}\epsilon$ cannot be Killing vector fields because, as we have seen, there are no everywhere future directed timelike (or null) Killing vector fields on de-Sitter spacetime. In fact the solutions of the twistor equation satisfy

$$\nabla_{\mu}\epsilon = \pm\frac{i}{2}\gamma_{\mu}\epsilon. \tag{101}$$

A simple calculation reveals, however, that this equation implies that the causal vector fields K^{μ} are in fact *conformal* Killing vector fields.

The situation for Anti-de-Sitter spacetime is completely different. No such element exists for $SO(p+1,2)$ or its universal cover and one does indeed have positive energy representations. One has energy raising and lowering operators

$$[\hat{E}, \hat{M}_i^{\pm}] = \pm M_i^{\pm}, \tag{102}$$

with $\hat{M}_i^{\pm} = \hat{M}_{0\,i} \pm i\hat{M}_{p+2,i}$, which increase the eigenvalues E by one unit. Thus one finds at the Lie algebra level representations such that the Anti-de-Sitter energy operator has integer spaced eigenvalues:

$$\hat{M}_{p+2\,0}|E\rangle = E|E\rangle \tag{103}$$

with

$$E = E_0 + n, n = r, r+1, 1\ldots \tag{104}$$

with r a non negative integer. The fractional part E_0 of the energy is constant in each irreducible representation and labels "superselection sectors" [28]. If

$$E = \frac{p}{k} \tag{105}$$

with p and k relatively prime, then we are in fact on the k-fold cover of AdS_{p+2}. If E_0 is irrational, then we must be on the universal cover. Actually for bosonic fields derived from supergravity fields it turns out that E_0 vanishes. Thus we are *de facto* on AdS_{p+2}.

The Poincaré translations are generated by

$$\hat{P}_{\mu} = \frac{1}{2}(\hat{M}_{\mu\,p+1} + \hat{M}_{\mu\,p+2}). \tag{106}$$

The special conformal transformations are generated by

$$\hat{K}_\mu = \frac{1}{2}(\hat{M}_{\mu\, p+1} - \hat{M}_{\mu\, p+2}). \tag{107}$$

The dilatation D corresponds to boosts and is thus given by

$$\hat{D} = \hat{M}_{p+1\, p+2}. \tag{108}$$

The quantized energy operator is given by

$$\hat{E} = \hat{M}_{p+2\, 0} = \hat{P}^0 + \hat{K}^0. \tag{109}$$

Now $\hat{E}$, $\hat{D}$ and $\hat{M}_{0\, p+1}$ span an $sl(2;\mathbb{R})$ subalgebra. Thus energy and dilatations do not commute. Hence they cannot be simultaneously diagonalised.

The question of integrality, however, can be thrown onto the behaviour under the operator $\hat{\tilde{J}}$.

9.1 Non-commutative Coordinates

Of course the generators $\hat{p}_\mu = \hat{M}_{p+1\, \mu}$ may be thought of as $p+1$ non-commuting "translations" since

$$[\hat{p}_\mu, \hat{p}_\nu] = i\hat{M}_{\mu\nu}. \tag{110}$$

In view of the great current interest in noncommutative geometry it may be worthwhile recalling a very early attempt to extract noncommutative coordinates from the AdS_{p+2} algebra. The idea was to take $\hat{x}_\mu = \hat{M}_{p+2\, \mu}$ as the "coordinates conjugate to the translations". One has

$$[\hat{x}_\mu, \hat{x}_\nu] = -i\hat{M}_{\mu\nu} \tag{111}$$

and

$$[\hat{p}_\mu, \hat{x}_\nu] = i\eta_{\mu\nu}\hat{M}_{p+1\, p+2}. \tag{112}$$

In eigenstates of the operator $\hat{D} = \hat{M}_{p+1\, p+2}$ we seem to be able to extract a spacetime version of the Heisenberg algebra! However, we certainly do not get a central extension in this way. In retrospect this victory looks a trifle hollow, but it is clearly closely related at a formal algebraic level to the Heisenberg Horospheres described earlier. It may indicate how to incorporate these older speculative ideas into the M-theory framework. The reader is referred to [34] for a recent and possibly related discussion.

10 CFT & ESU à la Lüscher and Mack

These authors [18] start with a conformal field theory on Minkowski spacetime $\mathbb{E}^{p,1}$ and then Wick rotate with respect to a constant time hyperplane to Euclidean space $\mathbb{E}^{p+1}$. Because the theory is conformally invariant it is assumed to extend to the conformal one-point compactification S^{p+1} on which the conformal group $\mathrm{Conf}(p+1) \equiv SO(p+2,1)$ acts.

We recall that the k-point compactification of a complete Riemannian manfold $\{M, g\}$ is a smooth compact Riemannian manifold $\overline{M}$ with metric $\overline{g}$ such that $\overline{M} \setminus \{x_i\}$, where x_i, $i = 1, \ldots, k$ are the infinity points, is diffeomorphic to M and on M, $\overline{g} = \Omega^2 g$ and where Ω is a smooth function on $\overline{M}$ which vanishes at the points x_i as one over distance squared. Stereographic projection (certainly known to Ptolemy and probably as far back as Hipparchus around 150 BC) provides the compactification in the present case. In spherical coordinates the spherical metric is

$$ds^2 = d\omega^2 + \sin^2 \omega d\Omega^2_{p+1}. \tag{113}$$

If $r = \tan(\omega/2)$, this becomes

$$ds^2 = (1 + \cos\omega)^2 (dr^2 + r^2 d\Omega^2_{p+1}). \tag{114}$$

One therefore has $\Omega = (1 + \cos\omega)$ which does indeed vanish like the distance squared as one approaches the infinity point at $\omega = \pi$. There is no $\mathbb{Z}_2$ factor here because we may think of compactified conformally $\mathbb{E}^{p+1}$ as the set of future directed null rays through the origin of $\mathbb{E}^{p+1,1}$. The Euclidean special conformal transformations correspond to boots.

Lüscher and Mack assume that $SO(p=1,1)$ will act nicely on any "Euclidean" conformal field theory on S^{p+1} and moreover that it will satisfy a version of Osterwalder-Schrader positivity with respect to reflection in an equatorial p-sphere. The round metric may be written as

$$\sin^2 \chi (d\tau^2 + d\Omega^2_p), \tag{115}$$

where

$$d\tau = \frac{d\chi}{\sin\chi}. \tag{116}$$

The coordinate τ covers the two-point conformal decompactification of S^{p+1}, the metric product $S^p \times \mathbb{E}$. The Osterwalder-Schrader reflection map θ is given by

$$\theta : \tau \to -\tau \tag{117}$$

and the associated semigroup mapping the upper hemisphere $\tau > 0$ into itself is given by

$$\tau \to \tau + a, \tag{118}$$

with $a \in \mathbb{R}_+$.

The net result is that one Wick rotates back to the Einstein Static Universe ESU_{p+1} by setting

$$T = i\tau. \tag{119}$$

Using this data Lüscher and Mack are able to show that one may obtain a Lorentzian CFT defined on the Einstein Static Universe, $\mathrm{ESU}_{p+1} \equiv \mathbb{E}^{0,1} \times S^3$. There exists a quantum mechanical Hilbert space $\mathcal{H}_{\mathrm{qm}}$ for such CFTs on which the universal cover $\tilde{O}(p+1,2)$ acts. As we have seen, ESU_{p+1} is the universal cover of the conformal compactification of Minkowski spacetime. The obvious question is whether the theory so defined will descend to the conformal compactification $\overline{\mathbb{E}^{p,1}} \equiv \mathrm{ESU}_{p+1}/\tilde{J}$ itself or a k-fold cover.

The answer given by Lüscher and Mack is that in general this is not possible. The existence of nonintegral dimensions, with fractional parts unequal, means that the $(\widehat{\tilde{J}})^k$ does not act projectively (i.e. up to a phase) on $\mathcal{H}_{\mathrm{qm}}$ and therefore one *cannot* project onto the space of invariant states.

Of course for very special CFTs it is not excluded that such projections are possible, but this requires very special anomalous dimensions. It is perhaps worth remarking here that the Euclidean approach to quantum field theory on S^4 adopted by Lüsher and Mack is almost *identical* to that used when one considers quantum fluctuations around an S^p universe "born from nothing" in quantum cosmology, cf.[12]. For an example in 2-dimensional CFT see based on the Schottky double of a Riemann surface see [40].

10.1 Supersymmetric Boundary Conditions

These were first addressed by Breitenlohner and Freedman. They found, in the absence of gravity, that one had two choices. Subsequently Hawking showed that demanding that the supergravity fields satisfy the boundary conditions necessary to permit the existence of asymptotic Killing spinors giving rise to an asymptotic Anti-de-Sitter superalgebra fixed this ambiguity uniquely. These boundary conditions are essential for the positive mass theorem to work in asymptotically Anti-de-Sitter spacetimes. The boundary conditions imply, however, that the boundary is invariant under $SO(p+1,2)$. In particular the boundary conditions will enforce periodicity with the Anti-de-Sitter period.

Hawking's original work was in four spacetime dimensions, but he has recently generalised it to all relevant dimensions.

10.2 Singletons

One of the remarkable features of the representation theory of the Anti-de-Sitter groups are the singleton and doubleton representations and their supersymmetric extensions. Rather than being connected with quantum field theory in the bulk, they are associated with a conformal field theory on the boundary. The simplest example is a conformally invariant scalar field ψ. This occurs as the lowest component of a superfield and has been interpreted as giving the transverse oscillations of the p-brane [17,14].

The equation of motion is

$$-\nabla^2\psi + \frac{p-1}{4p}R\psi = 0, \tag{120}$$

where R is the Ricci scalar of $S^p \times S^1$.

A simple calculation leads to modes of the form

$$Y_l \exp(i(l + \frac{p-1}{2})T) \tag{121}$$

where l is a nonnegative integer and Y_l is a spherical harmonic on S^p which behaves as $(-1)^l$ under the antipodal map on S^p.

Thus the transverse mode satisfies

$$\psi(\tilde{J}x) = i^{p-1}\psi(x). \tag{122}$$

Thus for the D3-brane $p = 3$ and the oscilations are invariant under $\tilde{J}^2$, for the M5-brane $p = 5$ so under $\tilde{J}$ and for the M2-brane $p = 2$ under $\tilde{J}^4$. This fits in remarkably well with the geometric picture based on the spacetime geometry. It seems that, as far as branes are concerned, Heraclitus may have been right after all!

11 Finite Temperatures and Event Horizons with Exotic Topology

The idea of thermodynamic equilibrium presupposes the existence of a timelike Killing field[8], Hamiltonian or energy operator $\hat{H}$ and conjugate time variable t. One aim is to compute the Gibb's partition function

$$Z(\beta; \mathcal{H}) = \mathrm{Tr}_{\mathcal{H}} \exp(-\beta\hat{H}), \tag{123}$$

where β and $\mathcal{H}_{\mathrm{qm}}$ is the quantum mechanical Hilbert space of the system one is considering.

It follows from the Heisenberg equations of motion and the commutativity or anticommutativity of fields at spacelike separations that the trace projects onto states which are periodic or antiperiodic in imaginary time $\tau = it$ with period β. This implies that correlation functions are also periodic or antiperiodic in imagainary time. An amusing example arises when one considers globally static coordinates in AdS_{p+2}. The finite temperature correlation functions are then periodic in both real and imaginary time. In the

[8] Strictly speaking, if only conformally invariant matter is considered, a timelike conformal Killing field may suffice. One may then, modulo conformal anomalies, pass to the conformally related stationary metric. This is important in cosmology, since all Friedman-Lemaitre-Robertson-Walker metrics are conformally static.

case of massless fields, when only poles are present, they may be expressed in terms of elliptic functions [52].

If additional mutually commuting conserved charges $\widehat{N}^i$ are involved, one introduces chemical potentials μ_i and considers

$$Z(\beta, \mu_i : \mathcal{H}) = \mathrm{Tr}_{\mathcal{H}} \exp(-\beta \hat{H} + \beta \mu_i \widehat{N}^i). \tag{124}$$

If the charges $\hat{H}$ and $\widehat{N}^i$ generate the Lie algebra g of a Lie group G, then $Z(\beta, \mu_i : \mathcal{H})$ is a sort of "character" in the representation of the semigroup element $\exp(-\beta \hat{H} + \beta \mu_i \widehat{N}^i)$ acting on Euclidean fields. In the case of spacetimes G is a maximally commuting subgroup of the isometry group and the charges $\widehat{N}^i$ are typically associated with angular momenta or Kaluza-Klein momenta. The chemical potentials μ^i are then interpreted as angular velocities or electrostatic potentials. The Wick rotation of the metric is slightly different in that case. Typically one anaytically continues to a complex section of the complexification $M_{\mathbb{C}}$.

11.1 Three Kinds of Static Metric

Depending upon which Killing field we take, we will get a different thermodynamics. Assuming that we maintain $SO(p)$-invariance, there are three natural (locally) static coordinate systems for AdS_{p+2}. The associated time translation is a one dimensional subgroup $G_1 \subset SO(2,1) \subset SO(p+1,2)$ acting on the coordinates say X^0, X^{p+1}, X^{p+2} and leaving invariant the coordinates X^i, $i = 1, \ldots, p$. The surfaces of constant time orthogonal to the timelines, i.e. to the orbits of G_1 in AdS_{p+1}, have the intrinsic geometry of hyperbolic space and are the intersections with the quadric of a one parameter family of hyperplanes passing through the origin acted upon by G_1.

The three possibilities correspond to the three conjugacy classes of one parameter subroups of $SO(2,1)$. They can be labelled by $k = 1, 0$ and are

- $SO(2)$ rotations in the $X^0 - X^{p+2}$ two-plane. The hyperplanes X^0/X^{p+2} = constant are always timelike. The system is globally static. There are no Killing horizons. Time translations correspond to $\hat{E} = \hat{M}_{p+2\,0} = \frac{1}{2}(\hat{P}^0 + \hat{K}^0)$. The metric is

$$ds^2 = -(1 + r^2)dt^2 + \frac{dr^2}{1 + r^2} + r^2 d\Omega^2_{p,1}, \tag{125}$$

  where $d\Omega^2_{p,1} = d\Omega^2_p$ is the metric on the unit p-sphere S^p.
- Null rotations. The hyperplanes $X^0/(X^{p+2} + X^{p+1})$ = constant are always timelike or null. The system is not globally static. There is an extreme Killing horizon at $r = 0$. Time translations correspond to $\hat{P}^0 = \hat{M}_{p+2\,0} + \hat{M}_{p+1\,0}$. The metric is

$$ds^2 = -r^2 dt^2 + \frac{dr^2}{r^2} + r^2 d\Omega^2_{p,0}, \tag{126}$$

where $d\Omega^2_{p,0}$ is the flat metric on $\mathbb{E}^p$.

- Boosts in the $X^0 - X^{p+1}$ two-plane. The hyperplanes $X^0/X^{p+1} = \text{constant}$ may be spacelike or timelike: the system is not globally static because there is a nondegenerate Killing horizon ar $r = 1$ with unit surface gravity. Time translations correspond to $\hat{M}_{p+1\,0} = \frac{1}{2}(\hat{P}^0 + \hat{K}^0)$. The metric is
$$ds^2 = -(r^2 - 1)dt^2 + \frac{dr^2}{r^2 - 1} + r^2 d\Omega^2_{p,-1}, \tag{127}$$
where $d\Omega^2_{p,-1}$ is the metric on hyperbolic space H^p.

It is of course possible to make identifications. For example one may convert $\mathbb{E}^p$ to a torus T^p and H^p to a closed hyperbolic manifold. In this way one obtains event horizons with exotic topologies. As stated above, this will lead to orbifold singularities if $k = 0$, which corresponds to horospheric coordinates with $z = 1/r$. Of course the relation of the coordinates (t, r) etc to the embedding coordinates is diifferent in all three cases.

These three examples can be used to define three kinds of (possibly locally) asymptotically Anti-de-Sitter boundary conditions with an associated concept of ADM mass. Taking out r^2 as a conformal factor, one sees that the conformal boundaries are the conformally flat manifolds:

- $S^p \times S^1$
- $\mathbb{E}^{p,1}$
- $H^p \times \mathbb{E}^{0,1}$.

In the last two cases these boundaries are geodesically complete as Lorentzian manifolds, but as conformal manifolds they are only subsets of the complete conformal boundary.

The cases $k = 1$ and $k = 0$ have no natural temperature, so it is possible to consider them at an arbitrary finite temperature $T = \beta^{-1}$. If $k = -1$, one must choose $\beta = 2\pi$. One may pass to imaginary time $\tau = it$ in the usual way and one gets the metric on hyperbolic space H^p which, in the cases $k = 1$ and $k = 0$, has been identified under the action of the integers generated by $\tau \to \tau + \beta$.

11.2 Tachyonic Black Holes

There are in addition black hole solutions, generalisations of the usual Kottler solution, of the form
$$ds^2 = -(r^2 + k + \frac{2M}{r^{p-1}})dt^2 + \frac{dr^2}{r^2 + k + \frac{2M}{r^{p-1}}} + r^2 \Omega^2_{p,k}. \tag{128}$$

The quantity M is proportional to the ADM mass. If $k = 1$, one finds that if the metric is to be nonsingular, in the sense that the singularuty at $r = 0$ is shielded by an event horizon, then M must be nonnegative. By contrast, if $k = -1$, negative values of M are allowed, as long as they are not too negative.

This fits in both with the AdS/CFT correspondence and with Wigner's observations. On the CFT side, in the case of three-branes, one finds that the Higgs fields of the $N = 4$ SUSY Yang-Mills theory have a coupling of the form:

$$-\frac{1}{12}\mathrm{Tr}R\Phi^2, \tag{129}$$

where R is the Ricci scalar of the boundary. In the case of $H^p \times \mathbb{E}^{0,1}$, this is negative and the coupling behaves like a tachyonic (i.e. negative mass squared) term. On the the group theory side, it is easy to see that the adjoint action of a rotation of π in the $X^0 - X^{p+2}$, that is an advance of of six Great Months, has the effect of reversing the sign of the relevant energy operator $\hat{M}_{p+1\,0}$.

These remarks also fit with some very old ideas about black holes in $p + 2$ dimensions in theories without a cosmological constant [42]. If the event horizon geometry is S^p rather than H^p, then the isometry group is $SO(p-1,1)\times\mathbb{R}$ rather than $SO(p)\times\mathbb{R}$. The latter is what Wigner called the little group, i.e. the stability group, of the timelike worldline of an ordinary particle. The latter the little group of the spacelike world line of a tachyon.

11.3 The Horowitz-Myers Conjecture

By reversing the role of one of the time and one of the spatial coordinates in the $k = 0$ case, Horowitz and Meyers find a black hole for which one of the spatial coordinates must be identified with period $\beta = \frac{4\pi}{p+1}(2M)^{\frac{1}{p+1}}$. This defines another boundary condition for which the conformal boundary is $S^1_\beta \times \mathbb{E}^{p-1} \times \mathbb{E}^{0,1}$. One may also identify points on the $\mathbb{E}^{p-1}$ factor to get a torus T^{p-1}. The solution is globally static: it does not have an event horizon. The spatial sections have topology $\mathbb{R} \times T^{p-1}$. Let us call this the Horowitz-Meyer version of the Kasner-Kottler spacetime, HM_{p+2}.

One might have thought that HM_{p+2} is an "excitation" of the identified space $AdS_{p+1}/\mathbb{Z}$ where the $\mathbb{Z}$ action is $x^1 \to x^1 + \beta$ in horospheric coordinates. However, working out the ADM Mass with respect to $AdS_{p+1}/\mathbb{Z}$ using the methods of [44] they find it to be negative!

Thus they lead to conjecture that it is HM_{p+2} which is the true ground state with respect to these boundary conditions and that there is some generalisation of the positive mass theorem to this setting. This is especially intriguing because HM_{p+2} admits no Killing spinors, ie. it is not BPS.

12 Concluding Observations

Having set the global scene, I shall make some observations about the the origin of the AdS geometry.

12.1 Nonlinear Realisations and Spontaneous Symmetry Breaking

The group manifold viewpoint makes it in some sense almost obvious that in any problem in which some sort of spontaneous breaking of translation and dilatation invariance is involved one can expect to be working on AdS_{p+2}. One may identify the coordinates x^μ as the Nambu-Goldstone bosons associated with translation invariance and $\phi = \ln z$ as that associated with dilatation invariance.

To see how consider, to begin with, the the case of the breakdown of a conventional global symmetry group G to an unbroken subgroup H. A low-energy effective Lagrangian can be constructed from maps from the world-volume of a p-brane to G/H. This requires a G-invariant metric on G/H. One may then construct Noether currents and obtain "current algebras".

For the p-brane one includes in G the group of translations transverse to the brane, the other variables being interpreted as additional scalar fields. The standard case of quantum field theory occurs when one has no transverse coordinates. The low energy dynamics of a single soliton defined in $\mathbb{E}^d$ is a another special case with $p = 0$ except that one typically now has a, possibly curved, "moduli space" $\{M, g\}$ of classical solutions the coordinates of which include the positions of the soliton and perhaps some internal degrees of freedom, such as phases or scales. The moduli space will certainly admit the action of the Euclidean group $E(d)$ and the position coordinates are associated with the orbits in M of the translation subgroup. In the case of BPS solitons, one also has multimoduli spaces M_k decribing the motion of k solitons. They are not just the products $M\times^k$ of the single soliton moduli space but at large soliton separation often tend to a product, and thus include a copy of the configuration space $C_k(\mathbb{R}^d) \equiv (\mathbb{R}^d)^k/S_k$. As far as the low energy dynamics are concerned the solitions move in a nonrelativistic Newton-Cartan "spacetime" if the form $M_k \times \mathbb{E}^0$.

Now all this is very reminiscent of Helmhotz's operational ideas about the physical origin of the axioms of geometry. By geometry he of course meant noneuclidean space geometry. Being a nineteenth century physicist he not surprisingly based his ideas on the "free mobility of rigid bodies". In effect he regarded space as the coset of possible locations G/H, where G is a six-dimensional Lie group containing $H = SO(3)$ as the group of rotations of a rigid body about a fixed point. The possiblities then reduce to to the triple of symmetric Riemannian spaces with $G = (SO(4), E(3), SO(3,1))$. The first and last are of course related by symmetric space duality.

Had Helmholtz known about quantum mechanics he might have proceeded differently but arrived at the same result. He might have assumed the existence of a set of operators or observables the commutation relations of which generated the Lie Algebra g. He would then seek to realise them on some Hilbert space $\mathcal{H}_{rm}$. A simple way for him to do so would be to take $L^2(G/H, \mu_g)$, where μ_g is the Riemannian volume element with respect to the invariant metric on g. In this way noneuclidean geometry would arise naturally from quantum mechanical principles as a consequence of assumptions about physical systems. Obviously extra degrees of freedom could have been incorporated by passing to a bigger group G_{unifying}, the extra degrees of freedom being interpreted as higher dimensions.

To make this picture compatible with relativity and fit the real world is not easy because we have to incorporate a more sophisticated idea of time into the picture. However, some elements are clear. The obvious analogue of $SO(3)$ is $SO(3,1)$ and $SO(4)$, $E(3)$ and $SO(3,1)$ are replaced by $SO(4,1)$, $E(3,1)$ and $SO(3,2)$. We might begin by replacing quantum mechanics by quantum field theory.

One obvious point of difference with the nineteenth century viewpoint is that for many particles we have no simple analogue of multiparticle spacetimes. This is usually taken care of by second quantization in which everything is thought of as happening in the *same* spacetime. Of course one may always think of k-point bosonic correlation functions as being defined on the k-th symmetric power of spacetime, but the geometry is just given by the product metric, unlike the case of the BPS monopole moduli spaces, where it very definitely is not the product metric. Moreover to capture all the information, because it is usually inconsistent to confine attention to a definite number of particles, one would consider instead the disjoint union $\sqcup_k M^k/S_k$. There do exist covariant multitime formulations of the classical mechanics of k point particles interacting at a distance, but they have no single time, as opposed to multitime, Hamiltonian formulation and they have as yet resisted quantization.

12.2 Anti-de-Sitter Space as a Moduli Space

The idea of spacetimes as moduli spaces is in fact not new. Therefore, before discussing the application of these ideas to string theory, it may prove illuminating to recall some rather old ideas about "Sphere Geometry" which go go back to the nineteenth century in which de-Sitter spaces and their metrics arise naturally.

Consider, to begin with, the more familar case of spheres $S^d - 1$ in Euclidean space $\mathbb{E}^d$. This arises physically in sphere packing problems [54,55]. Spheres have the the equation

$$U\mathbf{x}^2 - 2\mathbf{x}.\mathbf{a} + V = 0. \tag{130}$$

The centre is at $\frac{\mathbf{a}}{U}$ and the radius $R = \sqrt{\frac{\mathbf{a}^2}{U^2} - \frac{V}{U}}$.

The $(d+2)$-tuple $a = (\mathbf{a}, U, V)$ and the $(d+2)$-tuple $\lambda a = (\lambda\mathbf{a}, \lambda U, \lambda V)$, $\lambda \neq 0$ give the same sphere. Moreover the radius will be real and nonvanishing as long as

$$\mathbf{a}^2 - UV > 0. \tag{131}$$

Thus the set of $d-1$ spheres in $\mathbb{E}^d$ corresponds to a subset of $\mathbb{RP}^{d+1}$. If we set $U = a^{p+2} + a^{p+1}$ and $V = a^{p+2} - a^{p+1}$, we will recognize the subset as the set of spacelike directions in $\mathbb{E}^{d=1,2}$, i.e. with de-Sitter spacetime identified under the antipodal map, $DeS_{d+1}/\mathbb{Z}_2$. In fact more can be said. We may make use of the freedom to rescale the coefficient U to set

$$R = U. \tag{132}$$

This means that $V = \frac{\mathbf{a}^2}{R} - R$ and hence a sphere a corresponds to the unit spacelike $(d+2)$-vector

$$a^A = (\frac{\mathbf{a}}{R}, \frac{1}{R}, \frac{\mathbf{a}^2}{R} - R). \tag{133}$$

Evidently the centre and radius $(\mathbf{a}, R)$ are horospheric coordinates for de-Sitter spacetime.

Now, if two spheres a and a' intersect, then the angle θ between them is given by

$$\cos\theta = \frac{1}{2RR'}\Big(R^2 + {R'}^2 - (\mathbf{a} - \mathbf{a}')^2\Big)^2. \tag{134}$$

Clearly

$$\cos\theta = a^A {a'}^B \eta_{AB} = \frac{1}{2}\Big(2 - (a^A - {a'}^A)^2\Big). \tag{135}$$

Thus the angle between to spheres, i.e. the conformal structure on the space of spheres, is encoded in the chordal distance, i.e. to the causal structure, of de-Sitter spacetime and vice versa. Under this correspondence, inversion in a sphere corresponds to reflection in the associated hyperplane. In this way sphere packing problems are related to discrete subgroups of $SO(p+1,1)$ generated by reflections [54,55]. Another application (if $d = 2$) is to the probablity distribution of craters on the moon. The metric of De-Sitter spacetime gives the "fractal", i.e. dilatationand translation invariant measure

$$\frac{d^d\mathbf{a}dR}{R^{d+1}}. \tag{136}$$

What about Anti-de-Sitter spacetime?. We have a similar picture but we must take care with signs. Consider a *spacelike* hyperbola of two sheets in Minkowski spacetime $\mathbb{E}^{p,1}$. Its equation is

$$Ux^2 - 2x^\mu a_\mu + V = 0. \tag{137}$$

The central spacetime event is at $\frac{a^\mu}{U}$ and we will get a two-sheeted hyperbola as long as

$$UV - a_\mu a^\mu > 0. \tag{138}$$

This corresponds to $AdS_{p+2}/\mathbb{Z}_2$. The "radius" is given by $\sqrt{\frac{\mathbf{a}^2}{U^2} - \frac{V}{U}}$. In other words the longest proper time between the two sheets is twice the " radius".

We may interpret the horospheric coordinates of Anti-de-Sitter spacetime as the coordinates of the central event and the size of the hyperbola. If we had chosen to consider the space of single sheeted hyperbolae in Minkowski-spacetime, we would have considered spacelike directions in $\mathbb{E}p, 1$ and arrived at a “spacetime” with two times.

12.3 Twistors and Line Geometry

The space of spheres or pseudospheres carries a natural conformal structure in all dimensions. The case of lines, however, in general will not. Plücker and Klein discovered that one may give a conformal structure to the space of lines in $\mathbb{RP}^3$. A line in $\mathbb{RP}^3$ determines up to scale and a simple bivector $\omega\Lambda^2(\mathbb{R})$. Two lines ω and ω' intersect if and only if $\omega \wedge \omega' = 0$. This quadratic form has signature $(3,3)$ and therefore the set of lines may be identified with the set of null rays in $\mathbb{E}^{3,3}$. This gives $(S^2 \times S^2)/\mathbb{Z}_2$ with metric of signature $(2,1)$[9]. Group-theoretically the projective group $PSL(3;\mathbb{R} \equiv SO(3,3)$. Thus if one is prepared to complexify one has a conflation of line geometry and and sphere geometry, that is of the projective geometry of three-dimensions and the conformal geometry of four-dimensions. This is closely related to Penrose’s Twistor programme. Any straight line in three dimensions may be lifted to a null geodesic in four-dimensional Minkowski spacetime. Penrose himself prefers to work over the complex, but one may restrict oneself to some real section and obtain some special cases.

12.4 Strings in Four Dimensions

With this set of ideas in mind it is instructive to consider a string theory in four spacetime dimensions. The Nambu-Goldstone modes include four spacetime coordinates. However, if dilatation symmetry is broken one should take the semidirect product G_5 of spacetime translations with the dilations. The extra Nambu-Goldstone mode is related to the Liouville mode of string theory. This naturally brings us to consider strings moving in G_5, i.e. one half of AdS_5. One might argue that the S^5 factor has to do with the Goldstone mode for an $SO(6)$ ”R” symmetry. This seems to be behind some of Polyakov’s thinking about Wilson loops which played an important role in suggesting the AdS/CFT correspondence.

The question the arises: from where do the extra generators come which are needed to take us behind the horizon? One possible answer, suggested to me by Tom Banks is as follows. It uses an old result from flat space CFT. Suppose that one has invariance under the Causality Group. Then one should have a canonical energy momentum tensor $T^{\mu}{}_{\nu}$ which is

[9] Because Majorana spinors play such a central role in supersymmetry it may sometimes be useful to recall that the space of projective Majorana spinors for four-dimensional Minkowski spacetime (with signature $(+++-)$ amy be identifed with $\mathbb{RP}^3$ [56].

- Conserved
$$\partial_\mu T^\mu{}_\nu = 0, \tag{139}$$
- Symmetric
$$\eta_{\mu\sigma} T^\sigma{}_\nu = \eta_{\nu\sigma} T^\sigma{}_\mu \tag{140}$$
and
- Trace-free
$$T^\mu{}_\mu = 0. \tag{141}$$

Then it follows that one has additional conserved currents coming from the additional conformal Killing vectors, K^μ, associated with special conformal transformations.
$$\partial(T^\mu{}_\nu K^\nu) = 0. \tag{142}$$

If the boundary conditions permit, one may be able to integrate these over a Cauchy surface to get the missing generators needed to extend the Causality group to the full conformal group. This is essentially the question which was addressed at a more rigorous level by Lüscher and Mack whose work was described above.

References

1. G W Gibbons and J B Hartle, Real tunneling geometries and the large-scale topology of the universe *Phys. Rev.* **D42** 2458-2468 (1990)
2. E Inönü and E P Wigner, On the Contraction of Groups and Their Representations *Proc Natl Acad Sci USA* **39** (1953) 510-1953
3. M Levy-Nahas, Deformations and contractions of Lie Algebras *J Math Phys* **8** (1967)1211-1222
4. W E Couch and R J Torrence, Conformal Invariance Under Spatial Inversion of Extreme Reissner Nordstrom Black Holes, *Gen Rel Grav* **16** (1984) 789-792
5. G W Gibbons, Aspects of Supergravity Theories, GIFT lectures San Feliu de Guixols, Spain (1984) in *Supersymmetry, Supergravity and Related topics* eds. F del Aguila, A Azcarraga & L E Ibanez (World Scientific, Singapore) (1985)
6. G C Hegerfeldt and G Hennig, Coupling of Spacetime and Internal Symmetry *Fort der Physik* **16** (1968) 491-543
7. A D Alexandrov, On Lorentz transformations, *Uspekhi Mat Nauk* **5** (1950) 187
8. E C Zeeman, Causality Implies Lorentz Invariance, *J Math Phys* **5** (1964) 490-493
9. W L Craig, *The Kalām Cosmological Argument* (Macmillan, 1979)
10. E R Harrison, *Cosmology* (Cambridge University Press, 1981)
11. H A Chamblin and G W Gibbons, Topology and Time Reversal, Proceedings of the Erice School on String Gravity and Physics at the Planck Scale, edited by N Sanchez and A Zichichi gr-qc/9510006
12. G W Gibbons and P Rychenkova, One-sided Domain Walls in M-theory
13. G W Gibbons, Branes as BIons, hep-the/98030203
14. G W Gibbons, Wrapping Branes in Space and Time, hep-th/9803206
15. L Castellani, A Ceresde, R D'Auria, S Ferrara, P Fré and M Trigiante, G/H M-branes and AdS_{p+2} Geometries, hep-th/9803039

16. M Awada, G W Gibbons and W T Shaw, , Conformal Supergravity, Twistors and the Super BMS Group *Annals of Physics* **171** 52–107 (1986)
17. G W Gibbons and P K Townsend, Vacuum interpolation in supergravity via super p-branes *Phys Rev Lett* **71** 3754-3757 (1994) hep-th/9307049
18. M Lüscher and G Mack, Global Conformal Invariance in Quantum Field Theory *Comm Math Phys* **41** (1975) 203-234
19. G Moore, Finite in All Directions, hep-th/9305139
20. A Zeghib, On closed anti de Sitter spacetimes, *Math Ann* **310** (1998) 695-716
21. G W Gibbons, The Elliptic Interpretation of Black Holes and Quantum Mechanics *Nucl Phys* **B271** 479-508 (1986)
22. G W Gibbons and H J Pohle, Complex Numbers, Quantum Mechanics and the Beginning of Time *Nucl Phys* **B 410** 117-142 (1993) gr-qc/9302002
23. G T Horowitz and D Marolf, A New approach to String Cosmology, hep-th/9805207
24. E Calabi and L Markus, Relativistic space forms, *Ann Math* **75** (1962) 63-76
25. R S Kulkarni, Proper Actions and Pseudo-Riemannian Space Forms , *Advances in Mathematics* **40** (1981) 10-51
26. H S Snyder, Quantized Space-Time *Phys Rev* **71** (1947) 38-41
27. H S Snyder, The Electromagnetic Field in Quantized Spacetime *Phys Rev* **71** (1940) 68-71
28. E P Wigner, Some Remarks on the Infinite de Sitter Space, *Proc Natl Acad Sci USA* **36** (1950) 184-188
29. T O Phillips and E P Wigner, de Sitter Space and Positive Energy *Group Theory and its Applications* ed E M Loebl Academic Press NY (1968) 631-676 New York (1968)
30. S W Hawking, C J Hunter and M M Taylor-Robinson, Rotation and the AdS/CFT correspondence, hep-th/9811056
31. J L Friedman, Lorentzian universes from nothing, *Class Quant Grav* **15** 2639-2644(1988)
32. L J Alty, The Generalized Gauss-Bonnet-Chern theorem *J Math Phys* **36** (1995) 3613-3618
33. G W Gibbons, G T Horowitz and P K Towsend, Higher-dimensional resolution of dilatonic black hole singularities, *Class Quant Grav* **12** (1995) 297-317
34. S Tanaka, Space-Time Quantization and Matrix Model hep-th/9808064
35. A Schild, Discrete Space-Time and Integral Lorentz Transformations *Phys Rev* **73** (1948) 414-415
36. G W Gibbons and S W Hawking, Cosmological Event Horizons, Thermodynamics and Particle Creation *Phys Rev* **D15** 2738-2751 (1977)
37. A Schild, Discrete Spacetime and Integral Lorentz Transformations, *Canadian J Math* **1**(1949) 29-47
38. Y Mimura and H Takeno, Wave Geometry, *Scientific Reports of the Research Institute For Theoretical Physics, Hiroshima University* **2** (1962)
39. G W Gibbons and A R Steif, Sphalerons and conformally compactified Minkowski spacetime *Phys Lett* **B346** (1995) 255-261
40. A Jaffe, S Klimek and A Lesniewski Representations of the Heisenberg Algebra on a Riemann Surface, *Comm Math Phys* **126** (1989) 421-431
41. G W Gibbons, C M Hull and N P Warner, The Stability of Gauged Supergravity *Nucl Phys* **B218** (1983) 173-190
42. G W Gibbons and D A Rasheed, Dyson-Pairs and Zero-Mass Black Holes *Nucl Phys* **B 476** (1996) 515- hep-th/9604177

43. G W Gibbons and P J Ruback, Classical Gravitons and their Stability in Higher Dimensions *Phys Lett* **B171** (1986) 390-394
44. G T Horowitz and R Myers, The AdS/CFT correspondence and a new positive energy conjecture for general realtivity hep-th/9808079
45. Quantum Field Theory in Anti-de Sitter space-time, *Phys Rev* **D18** (1978) 3565
46. P Breitenlohner and D Z Freedman Stability in gauged extended supergravity *Ann Phys* **144** (1982) 249
47. S W Hawking, The boundary conditions for gauged supergravity, *Phys Lett* **126B** (1983) 175
48. L F Abbot and S Deser Stability of gravity with a cosmological constant, *Nucl Phys* **B195** (1982) 76
49. G W Gibbons and D L Wiltshire, Spacetime as a Membrane in Higher Dimensions *Nucl Phys* **B287** (1987) 717-742
50. H Nicolai, Representations of Supersymmetry in Anti-de-Sitter Space, in *Supersymmetry and Supergravity '84* eds B de Wit and P Van Nieuwenhuizen (World Scientific, Singapore, 1984)
51. L Mezincescu and P K Townsend,Stability at a Local Maximum in Higher Dimensional Anti-deSitter Space and Applications to Supergravity *Ann Phys* **160** (1985) 406-419
52. B Allen, A Folacci and G W Gibbons, Anti-de Sitter Space at Finite Temperature *Phys Lett* **B189** (1987) 304-310
53. A Chamblin, Existence of Majorana Fermions for M-Branes wrapped in space and time hep-th/9812134
54. H J Hermann, G Mantica and D Bessis, Space-Filling Bearings, *Phys Rev Lett* **65** (1990) 3223-3226
55. D Bessis, Generalized Appollonian Packings, *Comm Math Phys* **134** (1990) 293-319
56. G W Gibbons, The Kummer Configuration and the Geometry of Majorana Spinors in *Spinors, Twistors, Clifford Algebras and Quantum Deformations* eds. Z. Oziewicz et al. (Kluwer, Amsterdam, 1993)
57. E Cartan *Abh Math Sem Univ Hamburg* **11** (1936) 111-162
58. L K Hua, Harmonic Analysis of functions of several complex variables in the classical domains *Trans Amer Math Soc* **6** (1963)
59. M B Stenzel, Ricci-flat metrics on the complexification of a rank one symmetric space, *Manusripta Mathematica* **80** (19930 151-163
60. A Uhlmann, Remark on the future tube, *Acta Physica Polonica* **24** (1963) 293
61. A Uhlmann, The closure of Minkowski Space, *Acta Physica Polonica* **24** (1963) 295-296
62. G W Gibbons and A R Steif, Sphalerons and Conformally Compactified Minkowski Spacetime *Phys Lett* **B 346** 255-261 (1995) hep-th/9412210
63. B G Schmidt, A new definition of conformal and projective infinity of space-times, *Comm Math Phys* **36** (1974) 73-90
64. P Lounestou and A Springer, Möbius Transformations and Clifford Algebras of Euclidean and Anti-Euclidean Spaces, in *Deformations of Structures* L Lawrynowicz ed. (Kluwer, Amsterdam, 1989) 79-90

Black Holes and Wormholes in 2+1 Dimensions

Dieter Brill

University of Maryland, College Park, MD 20742, USA
and
Albert-Einstein-Institut, Potsdam, FRG

Abstract. Vacuum Einstein theory in three spacetime dimensions is locally trivial, but admits many solutions that are globally different, particularly if there is a negative cosmological constant. The classical theory of such locally "anti-de Sitter" spaces is treated in an elementary way, using visualizable models. Among the objects discussed are black holes, spaces with multiple black holes, their horizon structure, closed universes, and the topologies that are possible.

1 Introduction

On general grounds (2+1)-dimensional spacetime was long considered unlikely to support black holes, before such solutions were discovered [1]. Black holes were commonly conceived as places where the effects of gravity are large, surrounded by a region where these effects are asymptotically negligible. Another possible reason is the idea that black holes are "frozen gravitational waves" and therefore exist only in a context where the gravitational field can have independent degrees of freedom. In 2+1 dimensional Einstein theory — that is, Einstein's equations in a 3-dimensional space-time of signature $(-++)$ — the pure, sourceless gravitational field has no local degrees of freedom, because in three dimensions the Riemann tensor is given algebraically by the Einstein tensor, which in turn is algebraically determined by the Einstein field equations. If there is no matter source and no cosmological constant, the Riemann tensor vanishes and space-time is flat; if there is no matter but a cosmological constant Λ, the Riemann tensor is that of a space of constant curvature $\Lambda/3$. Thus gravity does not vary from place to place and it does not have any wave degrees of freedom. These were some of the reasons why the possibility of black holes was discounted, and the discovery of black hole solutions in 2+1 D spacetimes with a negative Λ came as such a surprise.

The existence of (2+1)-dimensional black holes of course does not alter the absence of gravitational waves in (2+1)-dimensional Einstein spaces, nor the lack of variation of their curvature. The curvature of spacetimes satisfying the sourceless Einstein equations with negative Λ is constant negative, and the local geometry in the asymptotic region does not differ from that near the black hole. Indeed, black hole solutions can be obtained from the standard,

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 143–179, 2000.

simply connected spacetime of constant negative curvature (anti-de Sitter space, AdS space for short) by forming its quotient space with a suitable group of isometries.[1] One of the criteria on the isometries is that the quotient space should not have any objectionable singularities. For example, if the group contains isometries of the rotation type, with a timelike set of fixed points, then the quotient space will have singularities of the conical kind. Such singularities can represent “point” particles, and the corresponding spacetime can be interpreted as an interesting and physically meaningful description of the dynamics of such particles [4]. However, we confine attention to solutions of the sourceless Einstein equations with negative cosmological constant — whether black holes or not — that are at least initially nonsingular. Therefore we exclude such particle-like solutions. (Likewise, we will not consider the interesting developments in lower-dimensional dilaton gravity, nor other matter fields [5].)

On the other hand, the group used to construct our quotient space may have isometries that are locally Lorentz boosts, with spacelike sets of fixed points. The corresponding singularities are of the non-Hausdorff “Misner” type [6]. If such a singularity does not occur on an initial spacelike surface, and is hidden behind an event horizon, then the spacetime can be acceptable as a representation of a black hole. Finally, the isometry may not have any fixed points but still lead to regions in the quotient space that are to be considered singular for physical reasons, and such regions may again be surrounded by an event horizon, yielding other types of black holes.

Thus the proper criterion characterizing a black hole in this context is not a region of large curvature or an infinite red shift (in typical representations of AdS space itself, where there is no black hole, there is an infinite red shift between the interior and the region near infinity), but existence of an event horizon. This in turn requires the existence of a suitable $\mathcal{I}$, whose neighborhood is a region in which “distant observers” can survive for an arbitrarily long time without hitting a singularity. That is, there have to be causal curves (the worldlines of these observers) that can be continued to infinite proper time. For example, Misner space itself — the quotient of Minkowski space by a Lorentz boost — does not satisfy this criterion in any dimension, because all timelike curves intersect the non-Hausdorff singularity in a finite proper time. Thus the case $\Lambda = 0$ does not yield any black holes. The same is true, for similar reason, in the case $\Lambda > 0$. However, for $\Lambda < 0$ there are worldlines along which asymptotic observers can survive forever even when spacelike singularities are present. Our black holes will then not be asymptotically flat [7], but asymptotically AdS. We will see (in Sect. 3) that the usual definition of black holes can be applied to these spacetimes, and even before we have come to this we will speak of them as black holes.

[1] It appears that all locally AdS spacetimes can be obtained in this way [2]. This is not so for positive curvature [3].

We can understand the difference between the cases $\Lambda \geq 0$ and $\Lambda < 0$ as as a consequence of the positive "relative acceleration" of spacelike geodesics in spaces of negative curvature. Spacelike geodesics reaching the asymptotically AdS region will increase their separation without limit. The fixed points of the identification that generates a black hole — that is, the "singularities" — lie along a spacelike geodesic. Consider a set of observers located initially further and further towards the asymptotic region and along another spacelike geodesic, which does not intersect the geodesic of fixed points. The timelike distance of an observer from the singularity will then eventually increase without limit, so a sufficiently far-out observer can survive for an arbitrarily long time.

We note in passing that *timelike* geodesics in spacetimes of constant negative curvature have the opposite property: they accelerate toward each other. Thus $\Lambda < 0$ corresponds to a universal "attractive" gravity, and a black hole in such a spacetime exerts this same attraction on test particles, as a black hole should.

The quotient of the AdS universe with the group generated by a single finite isometry that is without fixed points, at least on some initial spacelike surface, yields a single black hole, called a BTZ spacetime (for its discoverers, Bañados, Teitelboim and Zanelli [1]). As we will see, one can make further identifications in a BTZ spacetime, obtaining more complicated black holes, and this process can be repeated an arbitrary number of times. Although the isometries used for the identification cannot be entirely arbitrary, the variety of possibilities and of the resulting spacetimes is quite large. These spacetimes cannot be described by their metric in one or in a few simple coordinate systems, because many coordinate patches would be needed to cover their possibly complicated topology. In principle such a spacetime is of course defined, and all its physical properties are computable, once we know the structure of the AdS isometries that generate it. But such a presentation does not give an accessible and easily visualizable picture of the spacetime. Therefore we prefer to describe the spacetimes combinatorially, by "gluing together" pieces of AdS space. This view allows one to gain many important geometrical insights directly, without much algebra or analysis (even if a few of these geometrical constructions may resemble a tour de force).

In Sect. 2 we consider the simplest, time-symmetric case. Because the extrinsic curvature of the surface of time-symmetry vanishes, this surface is itself a smooth two-dimensional Riemannian space of constant negative curvature. This class of spaces has been studied in considerable detail [8]. In particular, almost all two-dimensional spacelike topologies occur already within this class. Section 3 considers the time development of these spaces; we find that all the non-compact initial states develop into black holes. The horizon can be found explicitly, although its behavior can be quite complicated. Section 4 concerns spacetimes that are not time-symmetric but have angular momentum.

An important reason for studying the classical behavior of these spacetimes is their relative simplicity while still preserving many of the features of more realistic black hole spacetimes. They are therefore interesting models for testing the formalism of quantum gravity. We do not go into these developments but refer the reader to the recent book by S. Carlip [9].

2 Time-Symmetric Geometries

Three-dimensional AdS space has many totally geodesic ("time-symmetric") spacelike surfaces. Because the extrinsic curvature of such surfaces vanishes, they have constant negative curvature Λ. Each such surface remains invariant under a "little group" of AdS isometries, which are therefore isometries of the spacelike surface, and conversely each isometry of the spacelike surface can be extended to be an isometry of the whole AdS spacetime.[2] Therefore any identification obtained by isometries on the spacelike surface can likewise be extended to the whole spacetime. (AdS space identified by this extension coincides with the usual time development of the initial data via Einstein's equations where the latter is defined, but it even goes beyond any Cauchy horizon). Thus to identify the possible time-symmetric geometries it suffices to discuss the possible initial spacelike geometries — although this leaves the time development still to be made explicit.

2.1 Coordinates

Although most physically and mathematically interesting facts about constant negative curvature spaces can be phrased without reference to coordinates, and even usefully so, it is convenient for the elucidation and proof of these facts to have coordinates available. Because of the large number of symmetries of AdS spacetime, its geometry takes a simple form in a large number of coordinate systems, which do not usually cover all of the spacetime, but which exhibit explicitly one or several of these symmetries. The simplest coordinates are the redundant set of four X^μ, $\mu = 1, \dots 4$ in terms of which AdS space is usually defined, namely as an embedding in four-dimensional flat space with signature $(-,-,+,+)$ and metric

$$ds^2 = -dU^2 - dV^2 + dX^2 + dY^2 \tag{1}$$

by the surface

$$-U^2 - V^2 + X^2 + Y^2 = -\ell^2. \tag{2}$$

[2] Since AdS spacetime is an analytic continuation (both in signature and curvature) of the familiar spherical geometry, such properties can be considered extensions of the corresponding statements about spheres, mutatis mutandis for the difference in group structure, SO(4) vs SO(2,2). Analogous statements are true about surfaces of constant extrinsic curvature.

This spacetime is periodic in the timelike direction with the topology $S^1 \times R^2$; for example, for $X^2 + Y^2 < \ell^2$ the curves $(X, Y) = \text{const}$, $U^2 + V^2 = \ell^2 - X^2 - Y^2$ are closed timelike circles. In the following we assume that this periodicity has been removed by passing to the universal covering space with topology R^3, which we will call AdS space. If it is necessary to distinguish it from the space of Eq (2) we will call the latter "periodic AdS space." Either spacetime is a solution of the vacuum Einstein equations with a negative cosmological constant $\Lambda = -1/\ell^2$.

Eq (2) shows that AdS space is a surface of constant distance from the origin in the metric (1). It therefore inherits from the embedding space all the isometries that leave the origin fixed, which form the SO(2,2) group. AdS space can be described by coordinates analogous to the usual spherical polar coordinates as in Eq (9), but of greater interest are coordinates related to isometries that leave a plane fixed, and whose orbits lie in the orthogonal plane. These have the nature of rotations if the plane is spacelike (or double-timelike, such as the (U, V) plane), and of Lorentz transformations if the plane is timelike. Isometries corresponding to orthogonal planes commute, and we can find coordinates that exhibit such pairs of isometries explicitly. If the isometries are rotations, the coordinates cover all of AdS space; if they are Lorentz transformations the corresponding coordinates are analogous to Rindler coordinates of flat space, and need to be analytically extended in the usual fashion to cover all of the spacetime.

For example, if we choose rotations by an angle θ in the (X, Y) plane and by an angle t/ℓ in the (U, V) plane, and specify the respective orbits on the AdS surface by

$$U^2 + V^2 = \ell^2 \cosh^2 \chi \qquad \text{and} \qquad X^2 + Y^2 = \ell^2 \sinh^2 \chi$$

(so that, for example, $U = -\ell \cosh \chi \cos \frac{t}{\ell}$, $V = \ell \cosh \chi \sin \frac{t}{\ell}$) we obtain the metric

$$ds^2 = -\cosh^2 \chi dt^2 + \ell^2 \left(d\chi^2 + \sinh^2 \chi d\theta^2\right). \tag{3}$$

In order to describe the universal covering space we have to allow t to range $-\infty < t < \infty$, whereas θ has its usual range, $0 \leq \theta < 2\pi$, and similarly $0 \leq \chi < \infty$. Except for the usual polar coordinate singularity at $\chi = 0$, these coordinates cover all of AdS space by a sequence of identical ("static") two-dimensional spacelike surfaces $t = \text{const}$ having a standard metric of spaces of constant negative curvature $-1/\ell^2$. Because $U = 0 = V$ does not occur on (2), shifts in the t coordinate are true translations, without fixed points. These coordinates define timelike sections ($\theta = \text{const}$) and spacelike sections ($t = \text{const}$) of AdS space. Each of these can be represented in a conformal diagram, shown in Fig. 1.

We can define a "radial" coordinate (which really measures the circumference of circles) by

$$r = \ell \sinh \chi .$$

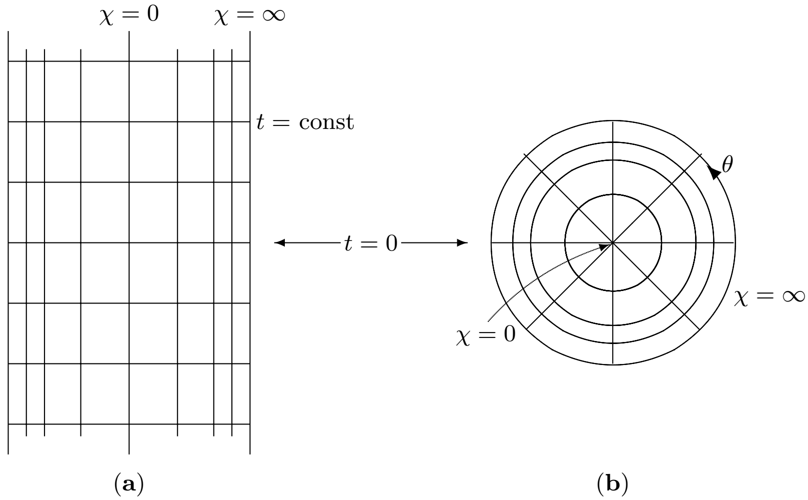


Fig. 1. Conformal diagrams of the static (or sausage) coordinates of Eq (3) in sections of AdS space. (**a**) The χ, t section, both sides of the origin. The right half is, for example, $\theta = 0$, and the left half, $\theta = \pi$. (**b**) The section $t = \text{const}$ is the 2D space of constant negative curvature, conformally represented as a Poincaré disk (see Sect. 2.2). The conformal factors are different in the two sections, so they do not represent sections of one three-dimensional conformal diagram. (For the latter see Fig. 4b)

The metric (3) then takes the form

$$ds^2 = -\left(\frac{r^2}{\ell^2} + 1\right) dt^2 + \left(\frac{r^2}{\ell^2} + 1\right)^{-1} dr^2 + r^2 d\theta^2 . \tag{4}$$

By choosing a different radial coordinate, namely

$$\rho = \ell \tanh \frac{1}{2}\chi$$

to replace the χ of Eq (3), we can make the conformally flat nature of the spacelike section explicit and keep the metric static:

$$ds^2 = -\left(\frac{1 + (\rho/\ell)^2}{1 - (\rho/\ell)^2}\right)^2 dt^2 + \frac{4}{\left(1 - (\rho/\ell)^2\right)^2} \left(d\rho^2 + \rho^2 d\theta^2\right) . \tag{5}$$

A picture like Fig. 1, with parts (**a**) and (**b**) put together into a 3-dimensional cylinder, can be considered a plot of AdS space in the cylindrical coordinates of Eq (5). Because of the cylindrical shape of this diagram these coordinates are sometimes called *sausage coordinates* [10]. Like the static coordinates of (3), these cover all of AdS space.

If we follow an analogous construction but use the timelike (X, U) and (Y, V) planes with orbits (in terms of a new coordinate χ)

$$-V^2 + X^2 = -\ell^2 \cosh^2 \chi \qquad \text{and} \qquad - U^2 + Y^2 = \ell^2 \sinh^2 \chi,$$

and new hyperbolic coordinates ϕ and t/ℓ, we obtain the metric

$$ds^2 = - \sinh^2 \chi dt^2 + \ell^2 \left(d\chi^2 + \cosh^2 \chi d\phi^2\right). \tag{6}$$

By defining

$$r = \ell \cosh \chi$$

we can change this to the Schwarzschild-coordinate form

$$ds^2 = - \left(\frac{r^2}{\ell^2} - 1\right) dt^2 + \left(\frac{r^2}{\ell^2} - 1\right)^{-1} dr^2 + r^2 d\phi^2, \tag{7}$$

which is usually derived from the "rotationally" symmetric ansatz — however, in this description of AdS space, ϕ has to be given the full range, $-\infty < \phi < \infty$ of a hyperbolic angle. The range of r for which the metric (7) is regular, $\ell < r < \infty$, describes only a part of AdS space, as can be seen from the explicit expression for the embedding in terms of these coordinates,

$$\begin{cases} U = \left(r^2 - \ell^2\right)^{1/2} \sinh \frac{t}{\ell} \\ V = r \cosh \phi \\ X = r \sinh \phi \\ Y = \left(r^2 - \ell^2\right)^{1/2} \cosh \frac{t}{\ell} \end{cases} \tag{8}$$

This regular region can be patched together in the usual way with the region $0 < r < \ell$ (Fig. 2), to describe a larger part of AdS space. But if it is desired (for whatever bizarre reason) to describe all of AdS space by analytic extensions of the coordinates (8), one needs also analytic extensions beyond the null surfaces $\phi = \pm\infty$ (or $r = 0$), which are quite analogous to the usual Schwarzschild-type "horizon" null surfaces $t = \pm\infty$ (or $r = \ell$). One then finds two disjoint regions of a third type (not shown in the figure because they extend perpendicular to the plane of Fig. 2a) in which r^2 is negative and ϕ is the timelike coordinate.[3]

Another interesting coordinate system is closely related to ordinary polar coordinates on the three-sphere:

$$\begin{cases} U = \ell \sin \frac{\tau}{\ell} \\ V = \left(r^2 - \ell^2\right)^{1/2} \cos \frac{\tau}{\ell} \\ X = r \cos \frac{\tau}{\ell} \cos \phi \\ Y = r \cos \frac{\tau}{\ell} \sin \phi \end{cases} \tag{9}$$

[3] Like all statements derived from embedding equations such as (8) this really applies to periodic AdS space, and should be repeated an infinite number of times for the covering AdS space itself. For example, there are an infinite number of regions of the three types in AdS space.

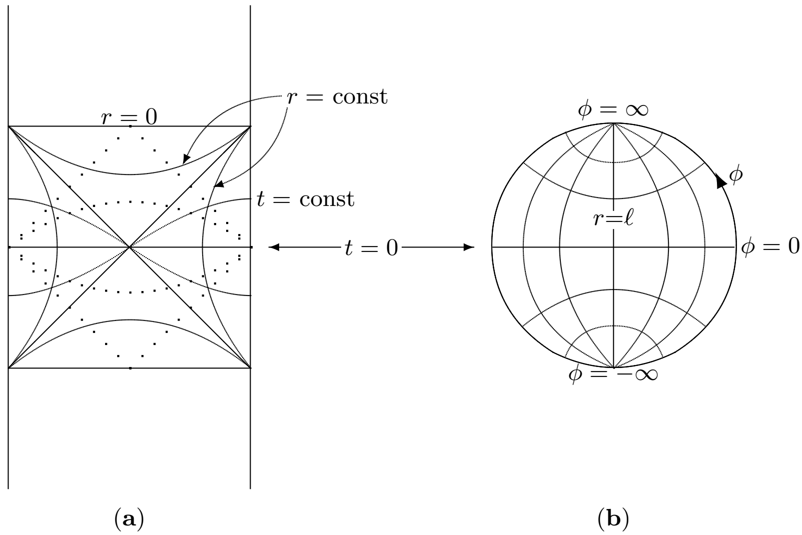


Fig. 2. Conformal diagrams of the "Schwarzschild" coordinates of Eq (8) in sections of AdS space. (**a**) An r, t section, continued across the $r = \ell$ coordinate singularity. The outer vertical lines correspond to $r = \infty$. The dotted curves show a few of the surfaces $\tau =$ const for the coordinates of Eq (10), with limits at $\tau = \pm\pi\ell/2$. (**b**) An r, ϕ section ($r > \ell$) is a two dimensional space of constant negative curvature, conformally represented as a Poincaré disk (see below). The approximately vertical curves are lines of constant r; they are equidistant in the hyperbolic metric. The approximately horizontal curves are lines of constant ϕ; they are geodesics in the hyperbolic metric. The outer circle corresponds to $r = \infty$

with the metric

$$ds^2 = -d\tau^2 + \cos^2\left(\frac{\tau}{\ell}\right)\left[\left(\frac{r^2}{\ell^2} - 1\right)^{-1} dr^2 + r^2 d\phi^2\right]. \tag{10}$$

This is a time development of the same initial data as in (7) (at $t = 0$ resp. $\tau = 0$) but with unit lapse function $N = 1$. The surfaces $\tau =$ const have constant extrinsic curvature, and they just cover the domain of dependence of those initial values.

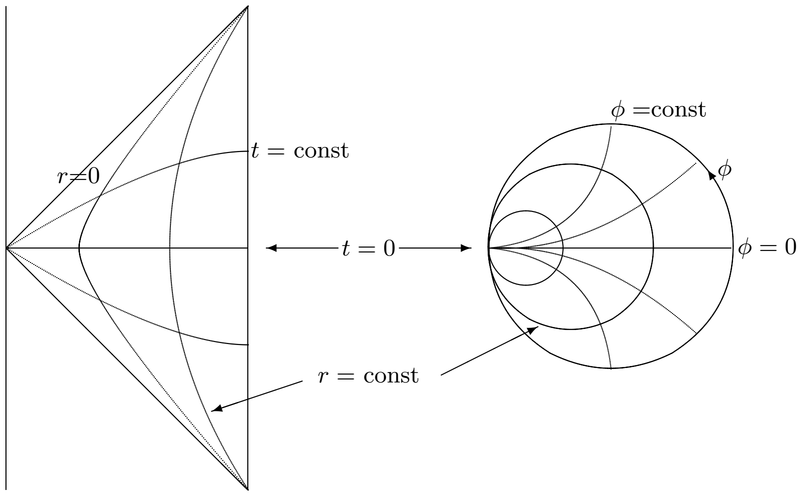
Finally one can introduce coordinates that correspond to the flat sections of de Sitter space:

$$\begin{cases} U + Y = r \\ U - Y = r\left(\phi^2 - t^2\right) + \frac{1}{r} \\ X = r\phi \\ V = rt \end{cases} \tag{11}$$

The metric then takes the form

$$ds^2 = -r^2 dt^2 + \frac{dr^2}{r^2} + r^2 d\phi^2. \tag{12}$$

Here the $r =$ const sections are manifestly flat.[4] Fig. 3 shows the conformal picture of these coordinates.



(**a**) (**b**)

Fig. 3. Conformal diagram of the "extremal" Schwarzschild coordinates of Eq (11) in sections of AdS space. (**a**) An r, t section. (**b**) An r, ϕ section. The lines $r =$ const are *horocycles* of the Poincaré disk.

The spacelike surfaces $t =$ const are conformally flat as are all two-dimensional surfaces, and as is manifest in Eq (5). Less trivially, the three-dimensional AdS spacetime also has this property, so neighborhoods of AdS space can be conformally mapped to flat space (one of the few cases where a three-dimensional conformal diagram exists). Such a map is the "stereographic" projection, a projection by straight lines in the embedding space from a point in the surface of Eq (2) onto a plane tangent to that surface at the antipodal point, analogous to the familiar stereographic projection of a sphere (Fig. 4a). By projection from the point $(U, V, X, Y) = (-\ell, 0, 0, 0)$ to the plane $U = \ell$ we obtain the coordinates (provided $U > -\ell$)

$$x^\mu = \frac{2\ell X^\mu}{U + \ell} \qquad X^\mu \neq U \tag{13}$$

[4] These subspaces are the analog in the case of Lorentzian metrics of *horospheres* of hyperbolic spaces (see, for example, [8]).

with the metric (where $X^0 = V$, $x^0 = t$)

$$ds^2 = \left(\frac{1}{1-r_c^2}\right)^2 \left(-dt^2 + dx^2 + dy^2\right) \quad \text{where} \quad r_c^2 = \frac{-t^2 + x^2 + y^2}{4\ell^2}\,. \quad (14)$$

This metric is time-symmetric about $t = 0$ but not static. It remains invariant under the Lorentz group of the flat 2+1-dimensional Minkowski space (t, x, y). In addition the origin may be shifted and the projection "centered" about any point in AdS space (by projecting from the corresponding antipodal point).

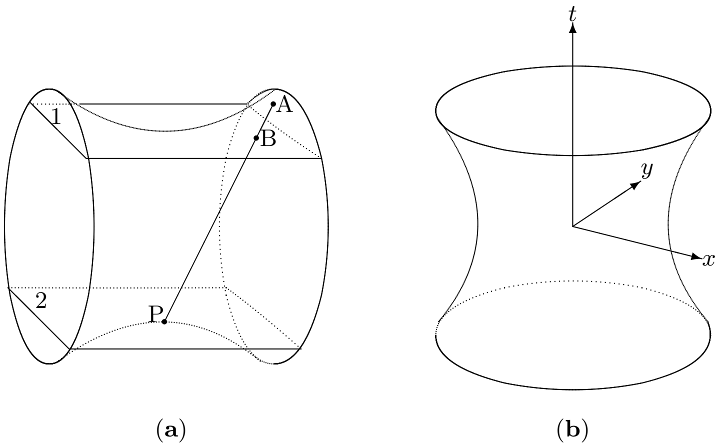


Fig. 4. AdS space in stereographic projection. (**a**) The hyperboloid is 2-dimensional AdS space embedded in 3-dimensional flat space as in Eq (2), restricted to $Y = 0$. It is projected from point P onto the plane 1 ($U = \ell$). The image of point A in the hyperboloid is point B in the plane. The part of the hyperboloid that lies below plane 2 is not covered by the stereographic coordinates. (**b**) When plotted in the stereographic coordinates (13), AdS space is the interior of a hyperboloid. The boundary of the hyperboloid is (part of) the conformal boundary of AdS space.

Because of the condition $U > -\ell$ the stereographic projection fails to cover a part of AdS space, even in the periodically identified version (Fig. 4a). The 3-dimensional conformal diagram is the interior of the hyperboloid $r_c = 1$, where the conformal factor of the metric (14) is finite (Fig. 4b). On the surface of time-symmetry, $t = 0$, the stereographic metric agrees with the sausage metric (5).

Many similar coordinate systems, illustrating various symmetries of AdS space, are possible; for examples see [11].

2.2 Isometries and Geodesics

To discuss the identifications that lead to time-symmetric black holes and other globally non-trivial 2+1-dimensional solutions we need a convenient representation of isometries and other geometrical relations in a spacelike initial surface of time-symmetry. Such a representation is the conformal map of Figs 1 and 2, in which this spacelike surface is shown as a disk, known as the *Poincaré disk*. This representation has been extensively studied (see, for example, [8]), and we only mention the features that are most important for our task.

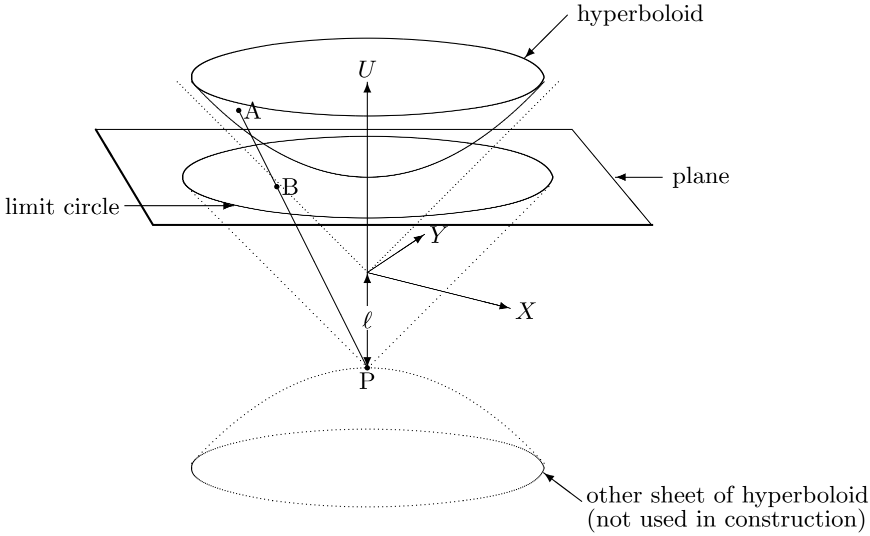


Fig. 5. The two-dimensional space H^2 of constant curvature $1/\ell^2$ is embedded in flat Minkowski space as one sheet of the hyperboloid of Eq (15). Under a stereographic projection from point P to the plane, point A on the hyperboloid is mapped to point B in the plane. Thus the hyperboloid (H^2) is mapped onto the Poincaré disk, the interior of the curve marked "limit circle"

All totally geodesic, time-symmetric surfaces H^2 in AdS space are isometric to the typical hyperboloid (Fig. 5) obtained by restricting Eq (2) to $V = 0$,

$$X^2 + Y^2 - U^2 = -\ell^2 \tag{15}$$

This surface has zero extrinsic curvature and therefore constant negative Gaussian curvature $-1/\ell^2$. The Poincaré disk can be obtained as a map of H^2 by the stereographic projection of Fig. 5, which illustrates Eq (13) when restricted to $V = 0$ similar to the way Fig. 4 illustrates it when restricted to $X = 0$. In this way all of H^2 is mapped into the interior of a disk of radius 2ℓ, whose boundary, called the limit circle, represents points at (projective or conformal) infinity. Because the map is conformal, angles are faithfully represented. Other geometrical objects in H^2 appear distorted in the Euclidean geometry of the disk, but by assigning new roles to these "distorted" objects and manipulating those according to Euclidean geometry one can perform constructions equivalent to those in the H^2-geometry directly on the Poincaré disk.

For example, on the surface H^2 as described by Eq (15), all geodesics are intersections of planes through the origin with the surface; that is, they satisfy a linear relation between X, Y, U. From Eq (13) it follows directly that Eq (15) becomes such a linear relation if x, y satisfy the equation of a circle that has radius $(a^2 - 4\ell^2)$ if it is centered at $(x, y) = (a_x, a_y)$, hence meets the limit circle at right angles. Because two such circles intersect in at most one point in the interior of the Poincaré disk, it follows that two geodesics in H^2 meet at most in one point (as in Euclidean space).

An important difference occurs if two geodesics do *not* meet: in Euclidean space they are then equidistant; whereas in the Poincaré disc the geodesic between points on two disjoint geodesics (Euclidean circles perpendicular to the limit circle) approaches a complete geodesic as the points approach the limit circle. Since the conformal factor in the metric of Eq (14), restricted to $t = 0$,

$$ds^2 = \left(1 - \frac{x^2 + y^2}{4\ell^2}\right)^{-2} \left(dx^2 + dy^2\right) \tag{16}$$

increases without limit as $x^2 + y^2 \to 4\ell^2$,on H^2 the geodesic distance between two given disjoint geodesics typically increases without bound as we go along the given geodesics in either direction. However, the geodesic distance between points on two given disjoint geodesics of course has a lower bound. If this is nonzero there is a unique geodesic segment of minimal length joining the two given geodesics at right angles to either.

On the other hand, if we have a family of equidistant curves, at most one of them can be a geodesic, and then the representation of the others on the Poincaré disk are arcs of circles, not perpendicular to the limit circle, but meeting the geodesic asymptotically at the limit circle. The curves $r =$ const of Fig. 2b are examples, with $r = \ell$ the geodesic of the family. These equidistant curves have constant acceleration (with respect to their arclength parameter), and they also illustrate how the conformal factor in (16) distorts the apparent (Euclidean) distances of the disk into the true distances of H^2.

Because the surface (15) in Minkowski space has constant extrinsic curvature, any isometry of the surface geometry can be extended to an isometry of the embedding space. But we know all those isometries: they form the homogeneous isochronous Lorentz group. Thus any Lorentz transformation implies, by the projection of Fig. 5, a corresponding transformation of the Poincaré disk that represents an isometry of H^2, and all H^2 isometries can be obtained in this way. In the Euclidean metric of the disk such transformations must be conformal transformations leaving the limit circle fixed, since they are isometries of the conformal metric (16).

Knowing this we can now classify[5] the isometries of H^2. Proper Lorentz transformations in 3D Minkowski space have an axis of fixed points that may be a spacelike, null, or timelike straight line. If the axis is timelike, it intersects the hyperboloid (15). If the axis is null, it intersects the hyperboloid asymptotically. If the axis is spacelike, it does not intersect the hyperboloid, but there are two fixed null directions perpendicular to the axis. Correspondingly on the Poincaré disk there is either one fixed point within the disk ("elliptic"), or one fixed point on the limit circle ("parabolic"), or two fixed points on the limit circle ("hyperbolic") for these transformations. Fig. 1b illustrates by the transformation $\theta \to \theta$+const the case with one finite fixed point (the origin). Figs. 2b and 3b illustrate by the transformation $\phi \to \phi$+const the case with two fixed points and one fixed point, respectively, on the limit circle ($\phi = \pm\infty$). In the case of two fixed points there is a unique geodesic ($r = \ell$ in Fig. 2b) left fixed by the isometry, and conversely the isometry, which we will call "along" the geodesic, is uniquely defined by the invariant geodesic and the distance by which a point moves along that geodesic.

Except for the rotation about the center of the disk as in Fig. 1b these are not isometries of the disk's flat, Euclidean metric, but they are of course conformal isometries of this metric. Such conformal transformations, mapping the limit circle into itself, are conveniently described as *Möbius transformations* of the complex coordinate

$$z = \frac{x+iy}{\ell} \quad \text{by} \quad z \to z' = \frac{az+b}{\bar{b}z+\bar{a}}\,, \tag{17}$$

where a, b are complex numbers with $|a|^2 - |b|^2 = 1$. When we consider an isometry or identification abstractly, it can always be implemented concretely by such a Möbius transformation. In particular, hyperbolic isometries are described by Möbius transformations with real a.

As the examples of Figs. 1-3 show, each of these isometries is part of a family depending on a continuous parameter (the constant in $\phi \to \phi$+const, for example). There is therefore an "infinitesimal" version of each isometry, described by a Killing vector ($\partial/\partial\phi$ in the example). Conversely an (orientation-preserving) isometry can be described as the exponential of its Killing vector.

2.3 Identifications

The hyperbolic transformations, which have no fixed points in H^2, are suitable for forming nonsingular quotient spaces that have the same local geometry as AdS space, and hence satisfy the same Einstein equations. In the context

[5] We confine attention to orientation-preserving transformations; they can be combined with a reflection about a geodesic (with an infinite number of fixed points) to obtain the rest.

of Fig. 2b and Eq (7) the transformation that comes to mind is described by $\phi \to \phi + 2\pi$. The quotient space is the space in which points connected by this transformation are regarded as identical, which is the same as the space in which ϕ is a periodic coordinate with the usual period. Eq (7) with this periodicity in ϕ already gives us the simplest BTZ metric for a single, non-rotating 2+1-dimensional black hole. It is asymptotically AdS, as shown by comparing Eqs (7) and (4).

The minimum distance between the two identified geodesics occurs at $r = \ell$ and is $2\pi\ell$. This is the minimum distance around the black hole, and plays the role of the horizon "area". If we identify ϕ with a different period $2\pi a$, we get a metric with a different horizon size. We can then redefine the coordinates so that ϕ has its usual period,

$$\phi \to a\phi, \qquad r \to r/a \qquad t \to at$$

and the metric takes this standard form, called the BTZ metric [1]:

$$ds^2 = -\left(\frac{r^2}{\ell^2} - m\right) dt^2 + \left(\frac{r^2}{\ell^2} - m\right)^{-1} dr^2 + r^2 d\phi^2, \tag{18}$$

where $m = 1/a^2$. Here the dimensionless quantity m is called the mass parameter. Although it can be measured in the asymptotic region, it is more directly related to the horizon size, the length of the minimal geodesic at the horizon, $2\pi\ell\sqrt{m}$.

The metric (18) is a solution also for $m = 0$, as shown by Eq (12), but that is not the AdS metric itself. The latter is also described by Eq (18), but with $m = -1$, as shown by Eq (4). By contrast, the $m = 0$ initial state is obtained by identifying the geodesics $\phi = 0$ and $\phi = 2\pi$ in Fig. 3b.

To describe the identification more explicitly, we may say that we have cut a strip from $\phi = 0$ to $\phi = 2\pi$ out of Fig. 2b, and glued the edges together. This strip is a "fundamental domain" for our identification, a region that contains images of its own points under the group only on its boundary, and that together with all its images covers the full AdS space. To obtain a fundamental domain for the BTZ black hole we might have used as the boundaries some other curve on the Poincaré disk and its image under the transformation, provided only that the curve and its image do not intersect. But since it is always possible to avoid apparent asymmetries by choosing boundaries composed of geodesics that meet at right angles, we will generally do so.

We can think of the identification in yet another way, by a process that has been called "doubling": cut a strip from $\phi = 0$ to $\phi = \pi$ from Fig. 2b, and cut another identical strip. Put one on top of the other and glue the two edges together, obtaining again the black hole initial state. The gluing makes the two strips reflections of each other with respect to either of the original edges. Back on the Poincaré disk the composition of the two reflections is a translation in ϕ by 2π, that is, the isometry of the identification.

Any (orientation-preserving) isometry of a hyperbolic space can be decomposed into two reflections [12]; hence any quotient space can be considered the double of a suitable region (possibly in another quotient space), and a fundamental domain is obtained from the region and one of its reflections.

The process of gluing together a constant negative curvature space from a fundamental domain of the Poincaré disk can be reversed: we cut the space by geodesics into its fundamental domain, make many copies of the domain, and put these down on the disk so that boundaries coming from the same cut touch, until the entire disk is covered. The resulting pattern is called a "tiling" of the disk (although the "tiles" corresponding to the $t = 0$ section of the BTZ black hole look more like strip flooring). Thus we have two equivalent ways of describing our identified space: by giving a fundamental domain and rules of gluing the boundaries, or by a tiling together with rules relating each tile to its neighbors.

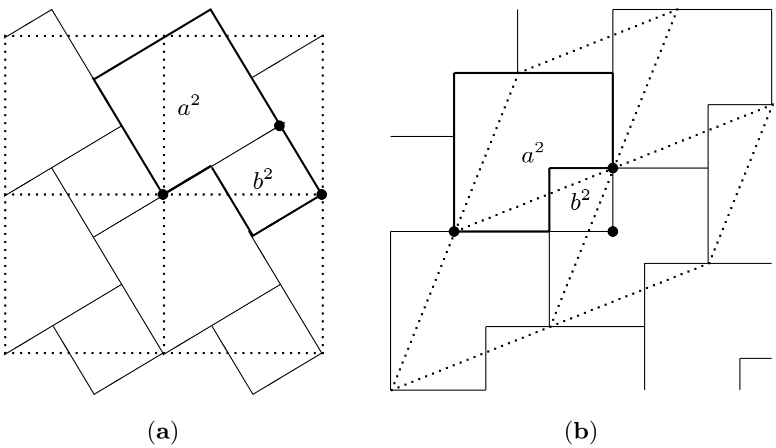


Fig. 6. Two different ways of tiling the plane prove the theorem of Pythagoras in (**a**) Euclidean space and (**b**) Minkowski space

Tiling and Pythagoras To fix ideas, consider an application of tiling found among the numerous proofs of the theorem of Pythagoras (a local boy who contributed to the early fame of Samos). This proof is based on the fact that all fundamental domains of a given group of isometries have equal area. In the Euclidean plane we consider the group generated by two translations specified in direction and amounts by two adjacent sides of the square above the hypotenuse of a right triangle, whose vertices are the three larger dots in Fig. 6a. This square is a fundamental domain of the group, and part of the tiling by this square is shown by the horizontal and vertical dotted lines. The region drawn in heavy outline is an alternative fundamental domain of the same group of isometries, and that domain is made from the squares above the sides of the same triangle. Part of its tiling of the plane is shown by the lightly drawn lines of Fig. 6a. Either fundamental domain can be glued together to form the same quotient space, a "square" torus, so the areas are equal, $c^2 = a^2 + b^2 =$ area of torus.

In special relativity the theorem of Pythagoras is valid with a different sign, $c^2 = a^2 - b^2$ if we choose the hypotenuse and one of the sides to be spacelike, and of course the right angles of the triangle and of squares are to be drawn in accordance with the Minkowski metric. Fig. 6b shows the proof by the tiling that derives from a Minkowski torus of area c^2. (Here we use, at least implicitly, the fact that the area of a two-dimensional figure is the same in Euclidean and Minkowski spaces if their metrics differ only by a sign.)

Embeddings To visualize the geometry of our glued-together surface — the $t = 0$ surface of a static BTZ black hole — it helps to embed this surface in a three-dimensional space in which the gluing can be actually carried out. This is analogous to the embedding of the $t = 0$ surface of the Schwarzschild black hole, with one angle suppressed, in three-dimensional flat space (the surface of rotation of the *Flamm parabola* [13]). For the BTZ initial surface only a finite part can be so embedded. The embedding stops where the rate of increase of circumference of the circle $r =$ const with respect to the true distance in the radial direction exceeds that rate in flat space. (The remainder of the surface could then be embedded in Minkowski space, but the switch between embeddings is an artifact and corresponds to no local intrinsic property.) However, the entire surface can be embedded in H^3, the *Riemannian* (positive definite metric) space of constant negative curvature. By the obvious generalization of the Poincaré disk this space can be conformally represented as a ball in three-dimensional flat space. Figure 7 shows this embedding, where the surface for $m > 0$ is seen to have two asymptotic sheets, similar to the corresponding Schwarzschild surface.

2.4 Multiple Black Holes

We saw that a single hyperbolic isometry (call it a) used as an identification to obtain an AdS initial state always yields a (single) BTZ black hole state, with horizon size and location depending on a. For other types of initial states we therefore need to use more than one such isometry, for example a and b. Assuring that there are no fixed points (which would lead to singularities of the quotient space) would then seem to be much more difficult: If we know that a has no fixed point, then the whole group consisting of powers a^n has no fixed points (except the identity, $n = 0$); but for the group generated by two isometries a, b we have to check that no "word" formed from these and their inverses, such as $ab^{-2}a^3b$ has fixed points. Although this may seem complicated, it is easy if we have a fundamental domain such that the isometry

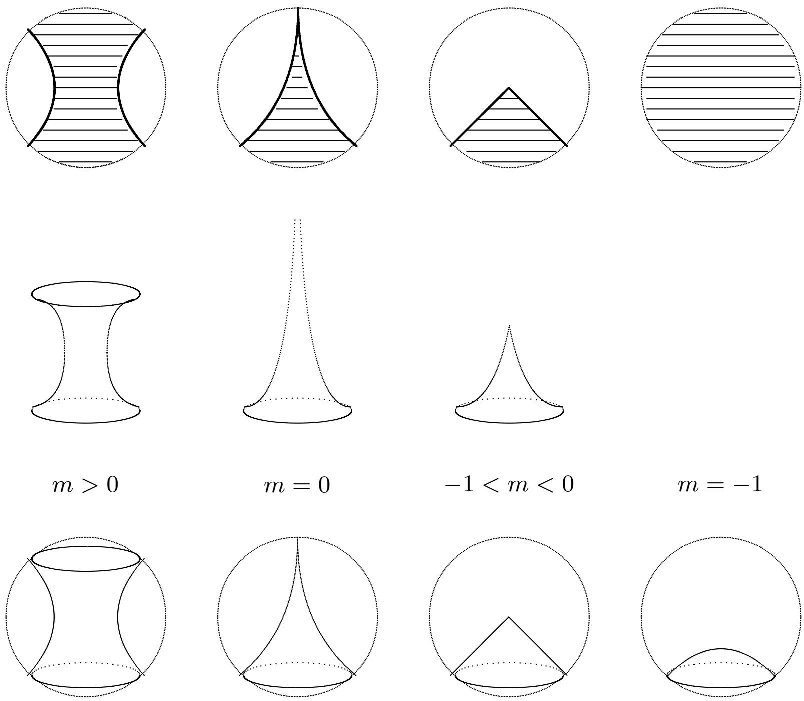


Fig. 7. Three representations of the geometry of the $t = 0$ geometry of metric (18) for different ranges of the mass parameter m: The BTZ black hole for $m > 0$; the extremal BTZ black hole for $m = 0$; the point particle (conical singularity) for $m < 0$; and AdS ("vacuum") space itself for $m = -1$. **Top row:** shaded regions of the Poincaré disk, to be identified in each figure along the left and right boundaries, drawn in thicker lines. **Second row:** an embedding of the central part ($r \leq \ell\sqrt{1-m}$) of these spaces as surfaces in three-dimensional flat Euclidean space. The embedding cannot be continued beyond the outer edges of each figure. **Bottom row:** the entire surface can be embedded in a 3D space of constant negative curvature, shown as a Poincaré ball. (The figure is schematic only; for example, the angle at the conical tips ought to be the same in the second and last row, to represent the same surface)

a maps one of a pair of boundaries into the other, and the isometry b does the same for a different pair of boundaries. Now tile the Poincaré disk with copies of this fundamental domain (see Fig. 10 for an example). Once we fix the original tile (associated with the identity isometry), there is a one-to-one correspondence between tiles and words. Therefore every non-trivial word moves all points in the original tile to some different tile, and there can be no fixed points in the open disk.

How to obtain such a fundamental domain? A simple way is by doubling a region bounded by any number k of non-intersecting geodesics [14]. Fig. 8a shows this for the case $k = 3$. In Fig. 8b we see the fundamental domain. Half of it is the original (heavily outlined) region, shifted to the right so that the center of the Poincaré disk lies on the geodesic boundary of the region rather than at its center. The other half is the reflection of this original region across that geodesic boundary. Thus $2k - 2 = 4$ boundaries remain to be identified in pairs, as indicated in the figure for the top pair. To construct the isometry that moves one member of such a pair into the other we find the unique common normal geodesic H_2 (shown for the bottom pair), and its intersection with the limit circle; these intersections are the fixed points of one of the hyperbolic isometries that have this fundamental domain. For example, in Fig. 8b the isometry associated with H_2 moves one of the bottom boundaries into the other. Similarly we find $k-2$ other isometries, each of them associated with a common normal. After the identification are made these common normals are smooth closed geodesics that separate an asymptotically AdS region from the rest of the manifold. We call such curves horizons. In addition to the $k - 1$ horizons found this way there is another one, so there is a total of k horizons. The additional one can be found in the above way from a different fundamental domain, obtained by reflecting the original region about a different geodesic boundary, but it is more easily found from the doubling picture, as shown by the H_3 in Fig. 8a.

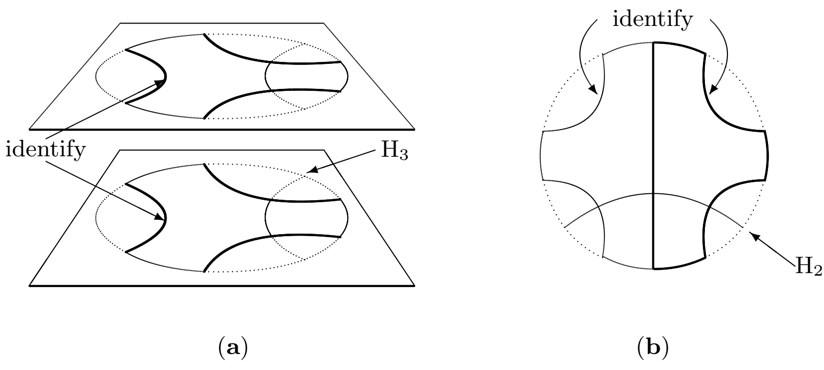


Fig. 8. Initial state for an AdS spacetime containing three black holes. (**a**) Representation by doubling a region on the Poincaré disk. The top and bottom surfaces are to be glued together along pairs of heavily drawn curves, such as the pair labeled "identify". The resulting topology is that of a pair of pants, with the waist and the legs flaring out to infinity at the limit circle. The heavier part of the curve H_3 becomes a closed geodesic at the narrowest part of a leg. (**b**) The fundamental region, obtained from one of the regions of part (**a**) by adding its reflection about the geodesic labeled "identify" in (**a**). In (**b**) only two boundaries remain to be identified; the top pair are so labeled. For the bottom pair the minimal connecting curve H_2 is shown

The topology of the resulting space may be easiest to see in the doubling picture: there are k asymptotically AdS regions, which can be regarded as k punctures ("pants' legs") on a 2-sphere. With each asymptotic region there is associated a horizon, namely the geodesic normal to the corresponding adjacent boundaries of the original region (because it is normal it will become a smooth, circular, minimal geodesic after the doubling). On the outside of each horizon the geometry is the same as that obtained from the isometry corresponding to that horizon alone, so it is exactly the exterior of a BTZ black hole geometry. Therefore the whole space contains k black holes, joined together inside each hole's horizon.

Parameters The time-symmetric (zero angular momentum) BTZ black hole in AdS space of a given cosmological constant is described by a single parameter, the mass m. For an initial state of several black holes we have analogously the several masses, and in addition the relative positions of the black holes. These are however not all independent. Consider a k black hole initial state obtained by doubling a simply-connected region bounded by k non-intersecting geodesics. Find the k minimal geodesic segments σ_i between adjacent geodesics.[6] The parts s_i of the original geodesics between the end-points of those segments, together with the segments σ_i themselves, form a geodesic $2k$-gon with right-angle corners. Clearly the σ_i are half the horizon size and hence a measure of the masses, and the s_i may be considered a measure of the distances between the black holes. If $2k-3$ of the sides of a $2k$-gon are given, then the geodesics that will form the $2k-2$ side (orthogonal at the end of the $2k-3$ side) and the $2k$ side (orthogonal at the end of the first side) are well-defined. They have a unique common normal geodesic that forms the $2k-1$ side, hence the whole polygon is uniquely defined. Thus only $2k-3$ of the $2k$ numbers measuring the masses and the distances of this type of multi-black-hole are independent. In the case $k = 3$ (corresponding to a geodesic hexagon) one can show that alternating sides (either the three masses or the three distances) can be *arbitrarily* chosen. Higher $2k$-gons can be divided by geodesics into hexagons, so at least all the masses (or all the distances) can be chosen arbitrarily. (The remaining $k-3$ parameters may have to satisfy inequalities.)

Composing the $2k$-gon out of geodesic hexagons means, for the doubled surface, that the multi-black-hole geometry is made out of $k-2$ three-black-hole geometries with $2k-6$ of the asymptotic AdS regions removed and the horizons glued together pairwise. In the five-black-hole example of Fig. 9 the three-black-hole parts are labeled 1, 2, and 3. One asymptotic AdS regions was removed from 1 and 3, and two such regions are missing from 2. The geometries obtained by doubling this are however not the most general time-

[6] Two geodesics of a set are adjacent if each has an end point (at infinity) such that between those end points there is no end point of any other geodesic of the set.

symmetric five-black-hole configuration. For example, in Fig. 9 the curve separating regions 2 and 3 is a closed geodesic. If we cut and re-glue after a hyperbolic isometry along this geodesic the geometry is still smooth; the operation amounts to rotating the top and bottom part of Fig. 9b with respect to each other, as indicated by the arrows. (In general we can make $k-3$ such re-identifications.) That the result is in general different after this rotation is shown, for example, by the change in angle between the boundary geodesic and another closed geodesic which, before the rotation, is indicated by the dotted line in Fig. 9a.

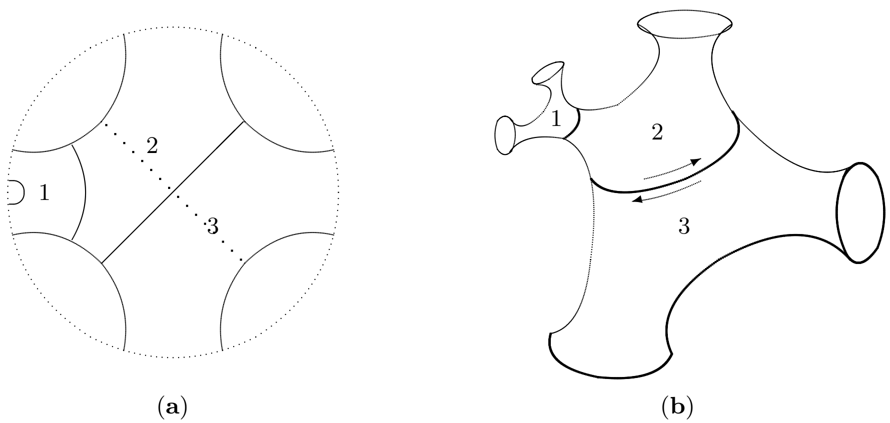


Fig. 9. A five-black-hole time-symmetric initial state is obtained by doubling the region on the Poincaré disk in (**a**). Part (**b**) shows a somewhat fanciful picture of the result of the doubling, cut off at the flare-outs, which should extend to infinity.

The $2k-3$ distance parameters and the $k-3$ rotation angles describe a $3k-6$-dimensional space of k-black-hole geometries. Equivalently we may say that a k-black-hole initial state is given by a fundamental domain bounded by $2k-2$ geodesics to be identified in pairs by $k-1$ Möbius transformations. Since each Möbius transformation depends on 3 parameters, and the whole fundamental domain can be moved by another Möbius transformation, the number of free parameters is $3k-6$. Such a space of geometries is known as a Teichmüller space, and the length and twist parameters are known as Fenchel-Nielsen coordinates on this space [8].

Instead of cutting and re-gluing along closed geodesics as in Fig. 9 one can do this operation on the identification geodesics used in the doubling procedure. For example, in Fig. 8a on the pair of geodesics marked "identify" one can identify each point on the bottom geodesic with one that is moved by a constant distance along the top geodesic. For the fundamental region this means the following: so far, whenever two identification geodesics on the boundary of the fundamental domain were to be identified, it was done by the unique hyperbolic transformation along the minimal normal geodesic between the identification geodesics. If we follow this transformation by a hyperbolic isometry along one of the identification geodesics, the two geodesics will still fit together, and the identified surface will be smooth but with a difference in global structure (like that produced by the re-gluing in Fig. 9). Of course the two transformations combine into one, and conversely any isometry that maps one identification geodesic into another can be decomposed into a "move" along the normal geodesic, and a "shift" along a identification geodesic. Since each hyperbolic transformation is a Lorentz transformation in the embedding picture (Fig. 5) the combination is again hyperbolic, so no finite fixed points (singularities) occur in this more general identification process.

If we identify with a non-zero shift, there is of course still a minimal geodesic between the two identified geodesics, but it is no longer orthogonal to those geodesics. Nevertheless the identified geometry is that of a black hole. To make the correspondence to the $\phi \to \phi + 2\pi$ identification of Eq (7) one would have to change the identification geodesics to be normal to the minimal one (which can complicate the fundamental domain).

Fixed points It is useful to understand the fixed points at infinity (the limiting circle of the Poincaré disk) of the identifications that glue a black hole geometry out of a fundamental domain of AdS space. The fixed points are directly related to the minimal geodesics associated with the identification, and they can indicate whether we have a black hole or not: there must be open sets free of fixed points if the initial data is to be asymptotically AdS. We know that the identifications can have some fixed points at infinity, but if the fixed points cover all of infinity, there is no place left for an asymptotically AdS region, and the space is not a black hole space. Thus even in the relatively simple time-symmetric case it is useful to understand the tiling and the fixed points of the Möbius transformationsassociated with the identifications.

As an example, consider again the three-black-hole case. Let a and b be the identifications of the top and the bottom pair of geodesics of a figure like 8b. Then the free group generated by these, that is, any "word" formed from a, b and their inverses A, B is also an identification. Since the identified geometry is everywhere smooth, none of these can have a fixed point in the finite part of the disk, so all fixed points must lie on the limit circle. The pattern of fixed points is characteristic of the identifications and constitutes a kind of hologram [15] of the multi-black-hole spacetime.

In Fig. 10 the initial fundamental domain is denoted by $\mathbf{1}$. The identifications are given by hyperbolic Möbius transformations a, b, with inverses A, B that connect the top and the bottom boundaries, respectively. Any "word" made up of these four letters is, first, also an identification. Secondly each word can be used to label a tile, because each tile is some image, $a\mathbf{1}$, $A\mathbf{1}$, $aB\mathbf{1}$,… of the initial domain $\mathbf{1}$, shown simply as a, A, aB, … in the figure. Finally, there is a closed minimal geodesic associated with each pair of

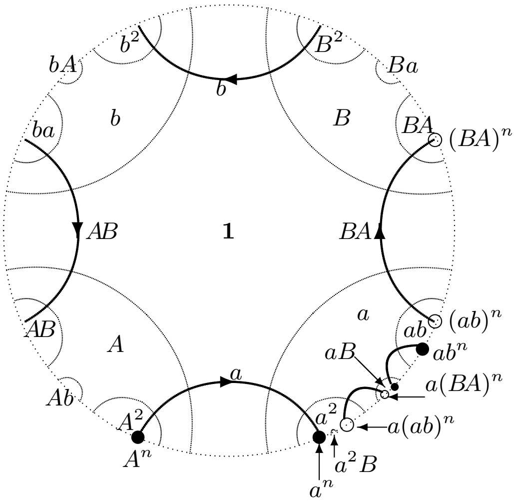


Fig. 10. Part of the tiling of the Poincaré disk obtained by "unwrapping" a three-black-hole initial geometry as in Fig. 8. A fundamental domain **1** is imaged by combinations of identification maps a and b and their inverses $A = a^{-1}, B = b^{-1}$. Repeating n times a map such as ab leads to a point $(ab)^n$ on the limit circle, in the limit $n \to \infty$. Some geodesics ("horizons") connecting such a limit point and its inverse limit (such as $a(ab)^n$ and $a(ab)^{-n} = a(BA)^n$) are shown as heavy curves

identified boundaries, hence each word also corresponds to a geodesic.[7] (For example, Ba connects $(Ab)^n$ to $(Ba)^n$.) Horizons are special geodesics that bound asymptotically AdS regions. Some of these are shown by the heavy curves. The ones that cut through the basic domain are labeled by the isometries that leave them invariant, a, b, and $AB = BA$. The words for the other horizons are obtained from these by conjugation, for example the horizon connecting the points labeled $a(ab)^n$ and $a(BA)^n$ is "called" by the word $a(BA)A$.

Every words is a hyperbolic isometry, hence has two fixed points on the Poincaré limit circle. We can find the fixed points by applying the word (or its inverse) many times to any finite region, because in the limit the images will converge to a point on the limit circle (see, for example, the equidistant curves in Fig. 2). Some of these fixed points are shown by open and by filled circles in the figure, and labeled by an nth power, where the limit $n \to \infty$ is understood. The two fixed points of a hyperbolic transformation define a geodesic that ends at them, and that is the minimal geodesic along which the transformation acts.

Because the infinity side of a horizon is isometric to the asymptotic region of a single black hole, there are no fixed points on that side of the horizon. (Cf. Fig. 2, where the only fixed points of the horizon isometry $\phi \to \phi +$ const are on the horizon $r = \ell$.) Between two different horizons (between open and filled circles of the figure) there will however be further horizons, with fixed points at their ends. Thus the set of fixed points for multi-black-holes has the fractal structure of a Cantor set.

[7] In this connection we regard a word, its inverse, and the permuted word as equal, in order to have a unique correspondence to geodesics; see [16].

By contrast, for some identifications the fixed points are everywhere dense on the limit circle. This happens, for example, if we try to build, by analogy to the multi-black-hole construction, a geometry containing three $m = 0$ black holes. The tiles, analogs of those of Fig. 10, would be "ideal" quadrilaterals, that is, each tile is a geodesic polygon whose four corners lie on the limit circle. This space is smooth and contains three ends of the type shown in the second column of Fig. 7 (instead of the "legs" in such pictures as Fig. 9); but since there is then no fixed-point-free region on the limit circle, this space is not asymptotically AdS and hence does not contain BTZ-type black holes.

2.5 Other Topologies

It is well known that time-symmetric AdS initial states, that is, spaces of constant negative curvature, admit a large variety of topologies. In the context of (orientable) black hole spaces one can construct all of these out of pieces of the three-black-hole space as in Fig. 7. These pieces are: three BTZ-exteriors, that is, the regions outside each of the three horizons; and one region interior to the horizons. The interior piece is sometimes called the "convex core" or "trousers."[8] Fig. 11 shows how other topologies can be constructed out of these pieces.

The resulting geometry is smooth if we choose the freely specifiable mass parameters of each exterior or core to match those of its neighbors at the connection horizons[9]: the intrinsic geometries then match, and the extrinsic geometries match because the horizons are geodesics. Conversely we can decompose a given k-black-hole initial geometry of genus g into asymptotic regions and trousers by cutting it along minimal circles of different and nontrivial homotopy types. We can choose $3g + 2k - 3$ such circles that divide the space into k exteriors $\mathcal{E}$ and $2g + k - 2$ trousers $\mathcal{T}$. Fig. 11 illustrates the construction for $k = 1$ and $g = 1$ (left) resp. $g = 2$ (right).

A fundamental region on the Poincaré disk, and hence the Möbius transformations that implement the identifications, can be constructed for these spaces in a similar way, by putting together geodesic, right-angle octagons representing trousers and analogous asymptotic regions. For example, the $k = 1$, $g = 1$ geometry can be represented by identifying two of the horizon geodesics of a trousers octagon and adding an exterior to the third. The resulting fundamental domain is bounded by geodesics, but it is not unique. We can cut it into pieces and re-assemble it in a different way [17], or we can cut the original space along some geodesics (not necessarily those of the trousers

[8] Previously we have used the image of flared pants' legs for the asymptotically AdS regions, which need to be cut off to obtain the core, so it would be more consistent to call the latter "cut-offs" or "shorts," but we will use "trousers."

[9] The horizons along which the legs were cut off from the cores may no longer be horizons of the space-time if the cores are re-assembled differently. Nonetheless, in the present section we will still call them by that name.

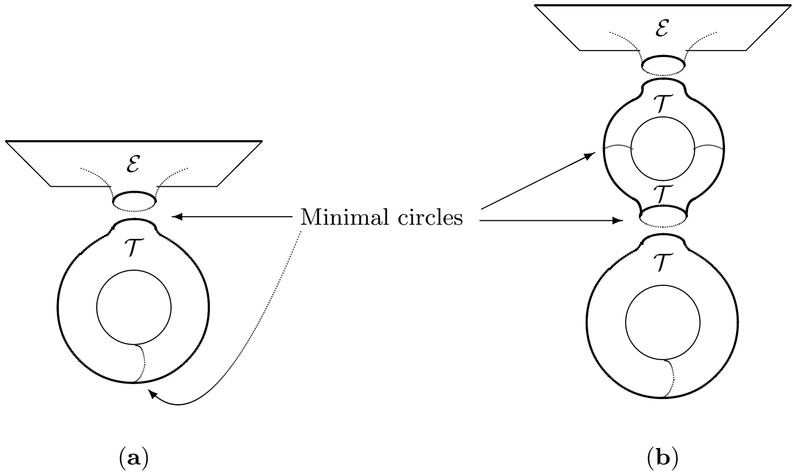
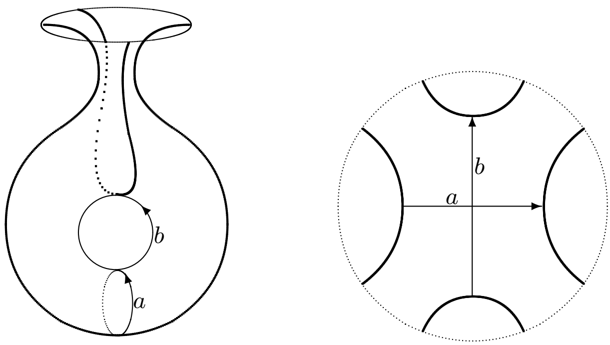


Fig. 11. Examples of the class of spaces considered here, constructed by sewing together one or several trousers and one asymptotic region. The latter looks asymptotically flat in this topological picture, but metrically it has constant negative curvature everywhere, just like the trousers

decomposition) only until it becomes one simply connected piece. If we can lay these geodesics so that they start and end at infinity and therefore do not cross we obtain a simple fundamental region bounded only by complete geodesics. Figure 12a shows the two geodesic cuts necessary for the case of our example of Fig. 11a, and the fundamental domain so obtained is seen in Fig. 12b. The pattern of tiling for this case is identical to that of Fig. 10, but the labeling is different. For example, rather than three horizon words there is only one, $abAB$, corresponding to the existence of only one horizon in the identified manifold.

Out of an even number of trousers only one can construct locally AdS initial data that contain no asymptotically AdS region at all. Such compact spaces can be interpreted as closed universes, and for lack of other physical content they can be considered to contain several black holes, associated with the horizons that were glued together in the construction. (Of course these horizons and black holes are only analogies, for example there are no observers that see them as black, i.e. of infinite redshift.)

Our reasoning about the number of parameters that specify a k-black-hole geometry can be generalized to the case that the internal geometry has genus g. If we cut off two asymptotic AdS regions and identify the two horizons that go with them, we decrease k by two and increase g by one. The number of parameters does not change: we lose one mass parameter, since the masses of the two horizons that will be identified have to be equal, but we gain one



(**a**) (**b**)

Fig. 12. Construction of a BTZ exterior with toroidal interior. Rather than cutting the geometry shown in part (**a**) by minimal geodesics as in Fig. 11, the cuts, shown by the heavy lines, are chosen to reach infinity and divide this into four regions. Thus one obtains the fundamental domain shown in part (**b**). The identifications a and b are shown as arrows. The lines of these arrows are also the minimal geodesics between the lines that are to be identified. In the identified manifold these are closed geodesics, as shown in part (**a**). The possible geometries are characterized by the lengths of these two closed geodesics, and the angle between them (shown here as 90°)

rotation parameter which specifies with what shift the horizons are to be identified. Thus from the formula in Sect. 2.4 we find that the (orientable) time-symmetric initial states, of genus g and k asymptotic AdS regions, form a $(6g+3k-6)$-dimensional Teichmüller space. If this number is non-positive, no state of that type is possible. (However, the formula cannot be applied to the time-symmetric BTZ initial state itself: it has one free parameter, the mass m, but no integral value of k makes the formula valid; the BTZ state is not a multi-black-hole geometry in the sense of this section.) For example, if we want a single exterior region ($k = 1$) we need a genus of at least $g = 1$ (Fig. 12). Here the number of parameters is $6g + 3k - 6 = 3$, for example the minimal distances (lengths of closed geodesics a and b) for each of the two identifications, and the angle between these geodesics. It is clear from the figure that these distances must be large enough, and the angle close enough to a right angle, that an asymptotic region remains in Fig. 12b. (If the geodesics crossed and formed a quadrilateral, there would be an angle deficit at the crossing point, which could be interpreted as a toroidal universe that is not empty, but contains one point particle.)

The formula for the number of free parameters tells us that there is no time-symmetric torus ($k = 0$, $g = 1$) initial state. However, all topologies of higher genus or with at least one asymptotic AdS region do occur; and the spatial torus topology does occur among all locally AdS spacetimes, for example as Eq (7) for $r^2 < \ell^2$ with ϕ and t periodically identified — the analog of a closed Kantowski-Sachs universe.

3 Time Development

The identifications used on a time-symmetric surface of AdS space to generate black hole and other initial values have a unique extension to all of AdS space, and thus define a unique time development (even beyond any Cauchy horizon). A fundamental domain in 3-dimensional AdS space can be generated by extending normal timelike geodesics from the geodesic boundaries of the two-dimensional fundamental domain on the initial surface. Due to the negative curvature of AdS space such *timelike* geodesics accelerate towards each other and will eventually cross. Such crossing of fundamental domain boundaries is the space-time analog of a conical singularity. A prototype of this is the "non-Hausdorff singularity" of Misner space [6]. Although not a curvature singularity, these points are considered not to be part of the spacetime. This in turn provides an end of $\mathcal{I}$ and hence the possibility of a black hole horizon.

A metric for the time development of the finite part of any multi-black-hole or multiply-connected time-symmetric initial geometry is provided by Eq (10) when we replace the expression in the bracket by the initial multi-black-hole metric. The result is a metric adapted to free-fall observers, and it shows that they all reach the singularity after the same proper time, $\tau = \pi\ell/2$, when the $\cos^2$ factor vanishes. (This can be seen geometrically from Fig. 4a, where geodesics are intersections with planes through the origin, and the collapse time is one quarter of the period around the hyperboloid.) But these coordinates do not cover the time development of conformal infinity (cf. the dotted curves in Fig. 2).

A more complete picture emerges from the continuation of the identification group to AdS spacetime, for example via the embedding of AdS space according to Eq (2). In the embedding of the initial surface in the 3-dimensional Minkowski space $V = 0$, each identification corresponds to a Lorentz "rotation" about some (spacelike) axis A. This is uniquely extended to an SO(2,2) "rotation" of the four-dimensional embedding space by requiring that the V-axis also remain invariant; that is, we rotate by the same hyperbolic angle about the A, V plane. This plane intersects the AdS space (2) in a spacelike geodesic of fixed points. All such geodesics from all the identifications are to be considered singularities after the identifications are made, so they are not points in the identified spacetime.

Three-dimensional pictures that include conformal infinity and all of the singularities can be had in sausage and in stereographic coordinates, Eqs (5) and (14). Because all timelike geodesics starting normally on a time-symmetric initial surface collapse together to a point C, all the totally geodesic boundaries of the fundamental domain also meet at C, forming a tent-like structure with a tip at C. Their intersections may be timelike or spacelike. If an intersection is timelike, the sides typically intersect there at a right angle "corner," and the intersection passes through the initial surface. If the initial geometry is smooth, such intersections are innocuous.[10] Spacelike intersections are called "folds" of the tent, and they are the geodesics of fixed points, which likewise meet at C.[11]

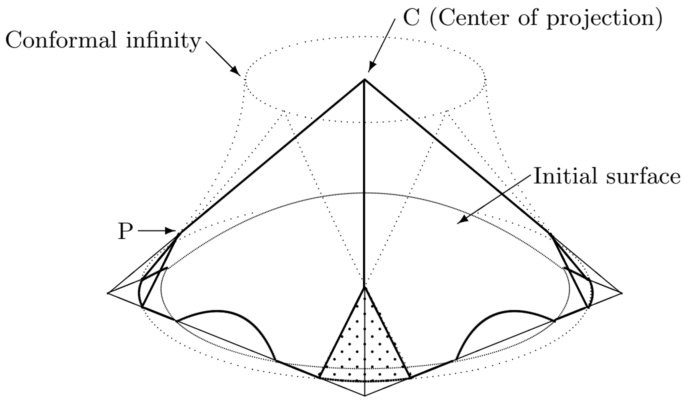


Fig. 13. The identification surfaces near the collapse point C in stereographic coordinates. AdS spacetime is the interior of the lightly dotted hyperboloid. The hyperboloid itself represents conformal infinity. The initial surface is a Minkowski hyperboloid (like that of Fig. 5) and in that sense is shown in its true metric. The triangular regions on the infinity hyperboloid, one of which is dotted, are the part of $\mathcal{I}$ that can be shown in this coordinate neighborhood

The tent has a simple, pyramid shape in a stereographic mapping centered at C. Since all geodesics through the center of the map are represented by straight lines in such a map, the sides of the tent are timelike planes (that is, linear spaces in stereographic coordinates), and the folds are spacelike straight lines. Figure 13 shows a tent with no corners but four folds. This can be the spacetime fundamental domain for the $k = 3$, $g = 0$ three-black-hole of Fig. 10 or for the $k = 1$, $g = 1$ toroidal black hole of Fig. 12, depending on the identification rule. In the three-black-hole case two of the folds, on opposite sides, are fixed points of the identifications a and b of Fig. 10 that

[10] We have not encountered such corners in our pictures, but they must appear in spaces composed only of trousers, for example in the time development of a $k = 0$, $g = 2$ surface that can be represented by a right-angled octagon on the Poincaré disk, as in Fig. 3b of [17].

[11] The reason that corners can be regular and but folds are not is that four corners can be put together to make a line without angle deficit, but no finite number of folds can eliminate the Misner-space singularity.

generate the group. The other two folds are fixed points of ab and of ba. For the toroidal black hole the fixed points of a and of b of Fig. 12 are not folds, they would be horizontal lines through the tip of the tent. Instead the folds are fixed points of $aba^{-1}b^{-1}$ and its three cyclic permutations. In each case the folds are fixed points of transformations associated with a horizon. All the other fixed points lie outside of the fundamental domain.

Because the stereographic picture is centered at a particular time, it can be misleading in that it does not exhibit the time symmetry about the initial surface, nor the early history before the time-symmetric moment. The time-independent sausage coordinates are more suitable for the global view of a black hole spacetime. Since the BTZ black hole (Fig. 14a) involves only one identification, its fundamental domain has only one geodesic of fixed points to the future of the initial surface, the $r = 0$ line in Fig. 2a. The sides of the tent are the surfaces $\phi = 0$ and $\phi = 2\pi$. Fig. 14b is the sausage coordinate version of Fig. 13.

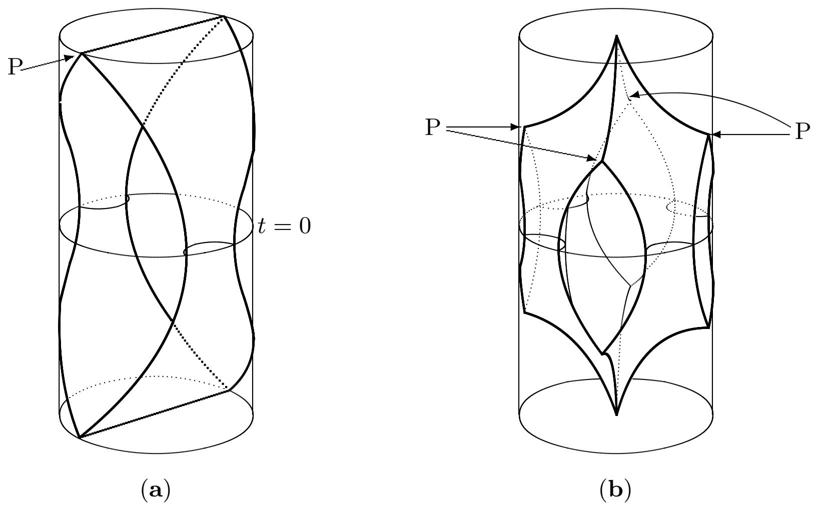


Fig. 14. Fundamental regions and their boundary "tents" in sausage coordinates, for (**a**) the BTZ black hole and (**b**) a three-black-hole or toroidal black hole configuration

Since folds are spacelike they extend to infinity, and therefore the initial fundamental domain must also have asymptotic regions. Conversely, the tent of an initial state without asymptotic regions has only corners and a tip but no folds: any closed time-symmetric AdS universe always collapses to a point in the finite time $\pi\ell/2$.

The holes in the tent are important for the black hole interpretation, for they are the regions at infinity, $\mathcal{I}$. The edges of the holes of course disappear once the identifications are made, and the only remaining boundaries of $\mathcal{I}$ appear as points such as those marked P in the figure. The backwards lightcone from P is the boundary of the past of $\mathcal{I}$, i.e. the horizon. It surrounds the singularity whose end is P. It is now clear that all the initial configurations that have a horizon in the sense of Sect. 2.4 do have spacetime horizons and hence are black holes: a horizon word extended to spacetime is an identification that has fixed points along some fold of the tent-shaped boundary of the spacetime fundamental domain. The intersection of the fold with conformal infinity is an endpoint of a $\mathcal{I}$, and the backwards lightcone of that endpoint is the spacetime horizon.

For a given fold we can consider a region in the fundamental domain sufficiently near infinity (spatially) and the fold (temporally) so that the only relevant identification is the one that has fixed points on that fold (because the other identifications would move points out of the region). In that region the spacetime is then indistinguishable from that of a BTZ black hole, and the spacetime horizon behaves in the same way as a BTZ black hole horizon. For example, the backward lightcone from P does intersect the initial surface in the minimal horizon geodesic. As we follow the horizon further backward in time it changes from the BTZ behavior only when it encounters other horizons or another part of itself, coming from another copy of the point P in the fundamental domain. For example, in the toroidal black hole interpretation of Fig. 14, all four openings of the tent are parts of one $\mathcal{I}$, and there is a single spacetime horizon consisting of the four "quarter" backwards lightcones from the four copies of the point P. As we go backwards in time below the initial surface these lightcones eventually touch and merge and shown in Fig. 15.

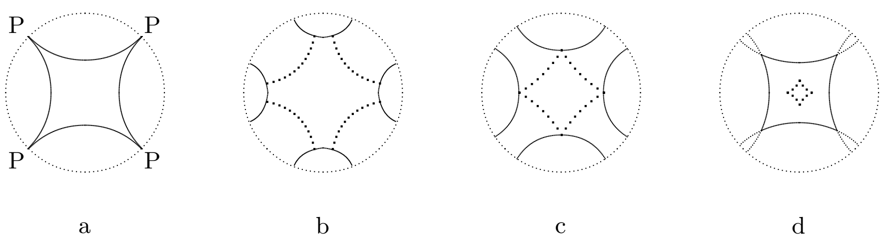


Fig. 15. Slices of the sausage in Fig. 14 to show the time development of the horizon. Part a is the latest and part d the earliest sausage time. The geometry of each time slice is the constant curvature space represented by a Poincaré disk. The geodesics shown by solid lines are to be identified as before for the toroidal black hole. Where these geodesics intersect we have a fixed point of some identification, a physical singularity. Slice a is at the sausage time of the end point P of $\mathcal{I}$. As we go backwards in time, the horizon (dotted arcs of circles) spreads out from those points at infinity. Slice b is the moment of time-symmetry. The horizon remains smooth until slice c, when its different parts meet each other at the identification surfaces. Prior to that time the event horizon has four kinks

3.1 Fixed Points at Infinity

As the above examples of multi-black-hole time developments show, any time-symmetric initial state with an asymptotic region ending at a horizon is isometric to a corresponding region of a BTZ black hole, so that each such region will look like a black hole from infinity for at least a finite time. It is maybe not so clear whether this is also true for the unlimited time necessary for a true black hole, for example because other singularities (fixed points) might intervene. By an interesting method due to Åminneborg, Bengtsson and Holst [18] one can directly find all of the universal covering space of $\mathcal{I}$ from a knowledge of spatial infinity on an initial surface. (The universal covering space gives information about horizons and is natural in many contexts, for example topological censorship questions reduce to existence of certain geodesics in AdS space [17].)

Since our black holes are quotient spaces of AdS space, the covering space of their $\mathcal{I}$ will be a subset of conformal infinity of AdS space. To describe this conformal infinity in a finite way we follow the usual Penrose procedure and multiply the AdS metric by a factor so that the resulting metric is finite in the asymptotic region. An obviously suitable conformal factor in Eq (5) is $(1 - (\rho/\ell)^2)^2$, giving the metric at infinity, $\rho = \ell$,

$$ds^2_\infty = 4(dt^2 + \ell^2 d\theta^2)$$

This is the flat metric of a cylinder of radius ℓ.

Consider first the covering space of $\mathcal{I}$ of a single black hole in this description, and recall that the identification is a "Lorentz boost" in the embedding space (2). As we apply this transformation n times to get the nth tile of the covering space, we are boosting the fundamental domain to the limiting velocity, and since the identification boundaries are timelike in our description, they become two null surfaces in the limit $n \to \pm\infty$. These null surfaces (called "singularity surfaces" in [19]) are then of course invariant under the identification transformation. Hence, if $\mathbf{K} = \partial/\partial\phi$ is the Killing vector corresponding to the identification, these surfaces are described by $\mathbf{K}^2 = 0$. They intersect where the vector $\mathbf{K}$ itself vanishes, that is at the fixed points at infinity on the initial surface and at the singularity inside the black hole.

The intersection of these surfaces with conformal infinity of AdS space is the boundary ($n \to \infty$) of the covering space of $\mathcal{I}$. To find it we only need to draw null lines from the endpoints of the horizon at $t = 0$ toward each other (Fig. 16a). Their future intersection is the nearest future fixed point to this initial surface, it is the end of $\mathcal{I}$, and the covering space of $\mathcal{I}$ is the diamond-shaped region between the future and past null lines. Furthermore the future null lines are also the intersection of the covering space of the horizon with conformal infinity, since the horizon is the backward null cone from the end of $\mathcal{I}$. Thus a knowledge of the initial fixed points gives us the "holographic" information about the exterior and the horizon of the black hole.

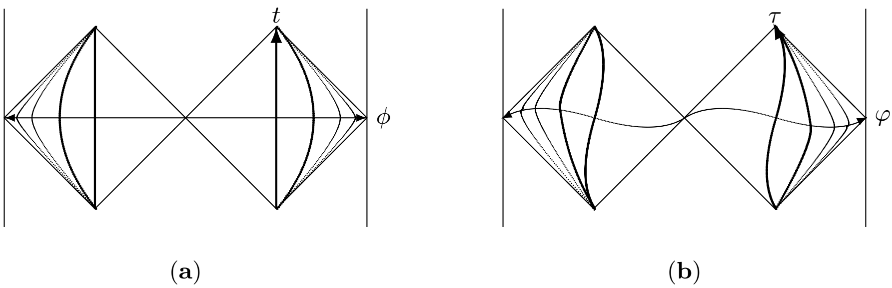


Fig. 16. Universal covering spaces of conformal infinity $\mathcal{I}$ for (**a**) non-rotating, (**b**) rotating black holes on the conformal infinity cylinder of AdS space. To show the cylinder in this flat picture it has been cut along the vertical lines, which are to be identified with each other in each part of the figure. The angle ϕ resp. φ runs from $-\infty$ to $+\infty$, in the direction of the arrow, in each of the diamond-shaped regions. The intersection of the fundamental domain with conformal infinity of AdS space is shown by the heavier boundaries. These boundaries are at the values 0 and 2π of ϕ resp. φ. A few of the tiles obtained by the isometries that change these angles by $2\pi n$ are shown for positive multiples n. In the limit $n \to \pm\infty$ the tiles converge to the null boundaries of the two diamond-shaped regions. These null boundaries intersect on the initial surface ($t = 0$ resp. $\tau = 0$) at the $n \to \infty$ limit at conformal infinity of the initial surface, and they end in the future at the end of $\mathcal{I}$

The situation for time-symmetric multi-black-holes is similar, except that the construction yields an infinite number of copies of the covering space of $\mathcal{I}$. We saw in Fig. 10 that each horizon word has two fixed points at spatial infinity. Each such pair yields a diamond-shaped region for which the transformation of its horizon word looks like that of Fig. 16a, and which is free of fixed points.

4 Angular Momentum

In the metric (18) for the static BTZ black hole, introduce new coordinates T, φ, R,

$$\begin{aligned} t &= T + \left(\frac{J}{2m}\right)\varphi \\ \phi &= \varphi + \left(\frac{J}{2m\ell^2}\right)T \\ R^2 &= r^2\left(1 - \frac{J^2}{4m^2\ell^2}\right) + \frac{J^2}{4m} \end{aligned} \tag{19}$$

where $J < 2m\ell$ is a constant with dimension of length, and define another new constant

$$M = m + \frac{J^2}{4m\ell^2}. \tag{20}$$

In terms of these new quantities the metric (18) may be written as

$$ds^2 = -N^2 dT^2 + N^{-2} dR^2 + R^2 \left(d\varphi + \frac{J}{2R^2} dT \right)^2 \tag{21}$$

where

$$N^2 = \left(\frac{R}{\ell} \right)^2 - M + \left(\frac{J}{2R} \right)^2 .$$

The metric (21) now looks like a (2+1)-dimensional analog of the metric for a black hole that carries angular momentum. Although metric (21) was obtained by a coordinate transformation from (18) and is therefore locally isometric to the latter (as all of our spaces are locally isometric to AdS space), it differs in its global structure: we have silently assumed that the new metric (21) is periodic with period 2π in the *new* angular variable φ, rather than in the old variable ϕ. This means, in coordinate-independent language, that we have changed the identification group that creates this new spacetime from AdS space. As for the non-rotating BTZ black hole, the new group for this "single" rotating black hole is still generated by all the powers of a single isometry of AdS space, but this isometry does not leave invariant a totally geodesic spacelike surface of time symmetry. The surface $T = \text{const}$ that is obviously left invariant by a displacement of φ is twisted, as measured by its extrinsic curvature, and this is one indication of the global difference from the static metric.

When only the one new coordinate φ changes by 2π, the old coordinates of (18) change by

$$t \to t + \frac{\pi J}{m} \qquad \phi \to \phi + 2\pi. \tag{22}$$

A change in either t or ϕ is of course an isometry of the metric (18), and because t and ϕ are coordinates, the two changes commute. The identification for a rotating black hole involves the two isometries applied simultaneously. Either one is a "boost" about an axis of fixed points; the change in ϕ has fixed points in the future, at $r = 0$, and the change in t has fixed points at $r = \ell\sqrt{m}$, the horizon of (18).[12] The combination of the two does not have any fixed points at all (either one moves points on the fixed axis of the other in the direction of the axis): the length R^2 of the corresponding Killing vector $\partial/\partial\varphi$ vanishes where the vector is null but not zero, since its scalar product with the finite $\partial/\partial T$ is the finite constant J. Earlier we argued that a spacelike set of fixed points of the identification isometry becomes a kind of singularity after the identification, and its removal from the spacetime gave us the end of $\mathcal{I}$ and associated horizon. What happens when we do not have this singularity?

[12] Since these isometries are also isometries of the periodically identified embedding (2), each axis of fixed points is really repeated an infinite number of times in AdS space itself.

4.1 Is It a Black Hole?

The geometry of metric (21) — more properly speaking, the geometry of its analytic extension, or of AdS space identified according to the $\varphi \to \varphi + 2\pi$ isometry exhibited by this metric — satisfies the definition of a black hole if we are somewhat creative about the definition of "singularity." We expect the singularity to occur at $R = 0$, but because there are no fixed points, the identified spacetime is regular there, and can be continued to negative R^2. But then the closed φ-direction becomes timelike, hence the spacetime has a region of closed timelike lines. We shall follow the usual practice to regard these as sufficiently unphysical that they should be eliminated from the spacetime, like a singularity. So we confine attention to $R > 0$.

Our spacetime then ends at the singularity surfaces where the square of the Killing vector $\partial/\partial\varphi$ vanishes, $R^2 = 0$. The corresponding r^2 of Eq (20) is negative. We recall from Sect. 2.1 that this occurs on two timelike surfaces in a region where ϕ is timelike, unlike the non-rotating black hole whose singularity occurs on the spacelike line $r = 0$. Since there is a singularity-free region between the two singularity surfaces, not all timelike lines that "fall into the black hole" (cross the horizon) end at the singularity; they can escape through the hole left open by the singularity surfaces, as is the case in a three-dimensional Kerr black hole. However, at conformal infinity the difference between R and r disappears, the two singularity surfaces come together at the point where the spacelike line $r = 0$ meets conformal infinity.

Thus the covering space of $\mathcal{I}$ for the rotating black hole looks the same as that of the non-rotating one that corresponds to it via Eq (20), only the identification is different, as shown in Fig. 16b. We see that $\mathcal{I}$ has an endpoint, there is a horizon, so the identified spacetime is a black hole.

We can recognize a (rotating) black hole in a spacetime by the presence of a closed, non-contractible spacelike geodesic γ. If we have such a γ we consider all spacelike geodesics that start normal to γ. We assume that these can be divided into two types, which we might call right-starting and left-starting (with respect to an arbitrarily chosen direction of γ). If all geodesics of one type reach infinity, then they cover the outside of a black hole. In this region the totally geodesic timelike surfaces normal to γ are surfaces of constant ϕ. Within these surfaces one can introduce coordinates so that the metric takes the form (21). (If the normals to those surfaces, not at γ, also integrate to closed curves after one circuit of γ, we have $J = 0$.)

4.2 Does It Rotate?

The asymptotically measurable properties of (2+1)-dimensional black holes can be defined in various way, for example: from the ADM form of the Einstein action; as the conserved quantities that go with the Killing vectors $\partial/\partial t$ and $\partial/\partial\varphi$; as the Noether charges associated with t- and φ-displacements; and so on [20,1]. All of these yield M as the mass and J as the angular momentum.

J can also be measured "quasi-locally" in the neighborhood of the horizon. We find an extremal closed spacelike geodesic (corresponding to $r = \ell$) and parallel transport an orthogonal vector around this geodesic. According to Eq (22) the hyperbolic "holonomy" angle between the original and rotated vectors is $\pi J/m\ell$.

4.3 Multiple Black Holes with Angular Momentum

It is fairly straightforward to extend the methods of Sect. 2.4 to obtain metrics with several asymptotic regions, or with non-standard topologies, that have angular momentum as measured in these asymptotic regions; the main difference is that we will deal with spacetimes rather than initial values. Our aim is only to show that rotating multi-black-holes are possible, and to indicate what the free parameters are.

We begin with a three-black-hole spacetime, whose time development can be described by the geometry of Fig. 13. We suppose that the front left and right surfaces are identified, and similarly the back left and right surfaces. The corresponding fixed points are the front and back edges of the pyramid. As we have seen, there is then a third black hole associated with the left and right edges (which are identified with each other). We cut this figure into two halves by the plane S (a totally geodesic surface) spanned by these left and right edges. This surface cuts the third black hole into two equal parts, which we can think of, for example, as $\phi = 0$ to $\phi = \pi$ and $\phi = \pi$ to $\phi = 2\pi$, respectively. Now we re-identify the two halves with a "boost" between them, that is an isometry with fixed points along the normal to the plane S at the center of the initial surface, as illustrated in Fig. 17.

The four planes stick out of the conformal infinity surface at the four bottom corners, uncovering four parts of conformal infinity. As in Fig. 13, the left and right infinity parts combine into one continuous region due to the identifications. So this spacetime has three conformal infinities with ends, and therefore represents three black holes.

The two black holes associated with the front and back edges, as seen from their respective asymptotic regions, are unchanged by this re-identification: by a "boost" isometry either of these edges and associated planes (but not both together) can be moved back to their old position. Since the two planes alone determine the asymptotic behavior of the black hole, either of these holes has the same mass, and vanishing angular momentum, as before. But the third black hole changes, because the left and right edges no longer lie in the same plane. As we go once around this third black hole, we cross the surface S twice, and its effects add (as a right-handed screw is right-handed from either end). The black hole therefore acquires angular momentum. Unfortunately this is not directly described by Eqs (19-22), because the "boost" in Eq (22) has fixed points at the horizon of the non-rotating black hole, whereas the fixed points of the boost of Fig. 17 lie along a geodesic connecting the asymptotic regions of the two other holes. However, for the third black

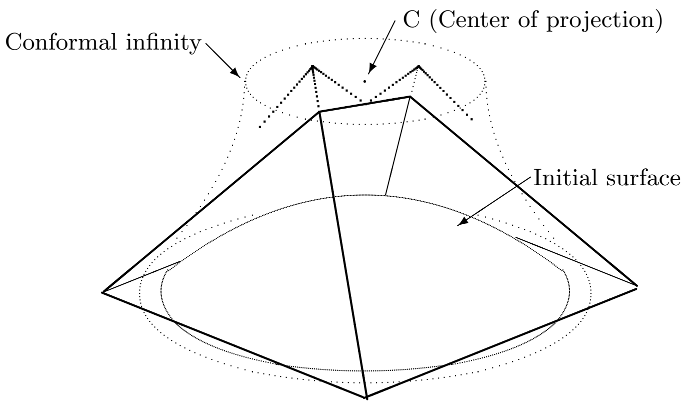


Fig. 17. A three-black-hole geometry obtained by cutting Fig. 13 into two tetrahedra by the plane S of the paper (passing through C), and re-gluing after an isometry with axis normal to that plane. The isometry moves the top of the front tetrahedron from C to the left (and up), and the top of the back tetrahedron from C to the right (and up). The dotted outlines show these two tops. The solid figure approximates the convex region bounded by the four planes that are identified pairwise (but it is *not* the fundamental domain). The edges where the planes intersect are drawn only to identify these planes; they are simplified as straight lines (but ought to be hyperbolic arcs, representing geodesics). Unlike in Fig. 13 the edges are not to be considered as singularities, except for the front and back edges, which are fixed points of the two basic identifications that generate this spacetime. The other "singularities" are the boundaries of the regions of closed timelike lines, not drawn (and not easily identified) in this figure.

hole this difference is asymptotically negligible: as seen from its own infinity it does have angular momentum. (Its standard form (21) would correspond to identification surfaces different from any of those drawn in Fig. 17.)

By a similar re-identification any one of a k-black-hole time-symmetric spacetime can be given angular momentum; further momentum parameters will be needed to describe how the asymptotic regions fit to an interior. Generally we expect one momentum parameter for each configuration parameter of the corresponding time-symmetric spacetime. For example, the toroidal black hole constructed as in Fig. 12 should allow three independent momenta. Of these the state in which there is angular momentum of the black hole as seen from infinity has been constructed [18]. (Another state with momentum can be obtained from Fig. 17 by identifying opposite rather than adjacent planes.)

5 Conclusions

We have seen that a considerable variety of black hole and multiply-connected spacetimes can be constructed by cutting a region out of anti-de Sitter space and identifying the cuts in various ways. Many of the properties, such as horizon structure and topological features of the time-symmetric spacetimes, have been investigated in detail. Comparatively little beyond existence is known about the spacetimes with angular momentum (but see [18]).

Acknowledgment

This research was supported in part by the National Science Foundation under Grant No. PHY94-07194.

References

For an extended list of references through 1994, see Carlip, S. (1995) The (2+1)-dimensional black hole. Class. Quant. Grav. **12**, 2853-2880

1. Bañados, M., Teitelboim, C., Zanelli, J., (1992) The black hole in three-dimensional space-time. Phys. Rev. Lett. **69**, 1849-1851; Bañados, M., Henneaux, M., Teitelboim, C., Zanelli, J. (1993) Geometry of the (2+1) black hole. Phys. Rev. D**48**, 1506-1525
2. Mess, G. (1990) Lorentz spacetimes of constant curvature. IHES preprint
3. Morrow-Jones, J., Witt, D. (1993) Inflationary initial data for generic spatial topology. Phys. Rev. D **48**, 2516-2528
4. See, for example, 't Hooft, G. (1993) Classical n particle cosmology in (2+1)-dimensions. Class. Quant. Grav. **10**, Suppl. 79-91. Such particles can form black holes: Matschull, H.-J. (1999) Black hole creation in 2+1 dimensions. Class. Quant. Grav. **16**, 1069-1095
5. See for example Strominger, A. (1994) Les Houches lectures on black holes. hep-th/9501071
6. Misner, C. (1967) Taub-NUT space as a counterexample to almost anything. In *Relativity Theory and Astrophysics I*, ed. J. Ehlers, Lectures in Applied Mathematics **8**, 160-169
7. For a recent exposition of asymptotically flat black holes see Heusler, M (1996) Black Hole Uniqueness Theorems. Cambridge University Press; for the asymptotically AdS case see [20]
8. Benedetti R., Petronio, C. (1992) Lectures on Hyperbolic Geometry. Springer, Berlin Heidelberg
9. Carlip, S (1998) Quantum gravity in 2+1 dimensions. Cambridge University Press
10. Holst, S. (1996) Gott time machines in anti-de Sitter space. Gen. Rel. Grav. **28**, 387-403; Åminneborg, S., Bengtsson, I., Holst, D., Peldán, P. (1996) Making anti-de Sitter black holes. Class. Quant. Grav. **13**, 2707-2714
11. Li, Li-Xin (1999) Time machines constructed from anti-de Sitter space. Phys. Rev. D**59**, 084016

12. Bachmann F. (1973) Aufbau der Geometrie aus dem Spiegelungsbegriff. Springer, Berlin Heidelberg
13. Flamm, L (1916) Beiträge zur Einsteinschen Gravitationstheorie. Phys. Z. **17**, 448-454, or see Misner, C., Thorne, K., Wheeler, J., (1973) Gravitation (p. 614). Freeman, San Francisco; Marolf, D. (1999) Space-time embedding diagrams for black holes. Gen. Rel. Grav. **31**, 919-944
14. Brill, D. (1996) Multi-black hole geometries in (2+1)-dimensional gravity. Phys. Rev. D **53**, 4133-4176; Steif, A. (1996) Time-Symmetric initial data for multi-body solutions in three dimensions. Phys. Rev. D **53**, 5527-5532
15. Susskind, L. (1995) The world as a hologram. J. Math. Phys. **36**, 6377-6396; Witten, E. (1998) Anti-de Sitter space and holography. Adv. Theor. Math. Phys. **2**, 253-291
16. Holst, S. (1997) On black hole safari in anti-de Sitter space. Filosofie Licentia-examen, Dept. of Phys., Stockholm Univ.
17. Åminneborg, S., Bengtsson, I., Brill, D., Holst, D., Peldán, P. (1998) Black holes and wormholes in 2+1 dimensions. Class. Quant. Grav. **15**, 627-644
18. Åminneborg, S., Bengtsson, I., Holst, S (1999) A spinning anti-de Sitter wormhole. Class. Quant. Grav. **16**, 363-382
19. Holst, S., Peldán, P. (1997) Black holes and causal structure in anti-de Sitter isomorphic spacetimes. Class. Quant. Grav. **14**, 3433-3452
20. Carlip, S., Gegenberg, J., Mann, R., (1995) Black holes in three-dimensional topological gravity. Phys. Rev. D **51**, 6854-6859; Abbott, L., Deser, S. (1982) Stability of gravity with a cosmological constant. Nucl. Phys. B **195**, 76; Brown, J., Creighton, J., Mann, R. (1994) Temperature, energy and heat capacity of asymptotically anti-de Sitter black holes. Phys. Rev. D **50**, 6394-6403

Open Inflation

S.W. Hawking

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Silver Street, Cambridge, CB3 9EW

In February this year, Neil Turok and I proposed a new model for inflation. The distinctive feature was that it produced an infinite, open universe, yet it satisfied the no boundary condition and came from an instanton of finite size, with a mass of the order of one gram. Our paper aroused a lot of interest, as shown by 37 citations on HEP-TH, but it brought a lot of opposition. This centered on three features of our model.

First, we were attacked for using the no boundary proposal. People like Linde and Vilenkin, claimed that one should use the quantum tunneling wave function instead.

Second, we were criticized because our instanton contained a singularity. It was said this was contrary to the spirit of the no boundary proposal; that the singularity would be naked and would make the universe non predictable.

Third, we invoked the anthropic principle, to avoid the model predicting a totally empty universe. We were attacked both for using anthropic arguments, and for the very low value for the density of the universe, that they seemed to lead to.

In this talk, I will describe the open inflation model that Neil and I proposed, and answer some of the objections that have been raised.

The original idea for inflation, was that in some way, the universe got trapped in what was called, a false vacuum state. A false vacuum state cor-

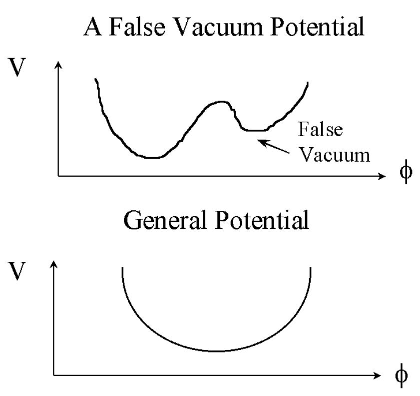


Fig. 1. False vacuum potential and general potential.

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 180–190, 2000.

responds to some scalar field being in a local minimum of the potential, that has more energy than the true vacuum, which is taken to have zero energy density.

$$T_{\mu\nu} = -\tfrac{1}{2} g_{\mu\nu} V(\phi) \tag{1}$$

False vacuum energy momentum tensor.

Because a false vacuum is Lorentz invariant, its energy momentum tensor must be proportional to the metric. Since the false vacuum has positive energy density, the coefficient of proportionality must be negative. The Einstein equations then imply that the scale factor increases exponentially with time.

In such a universe, the integral of one over the scale factor, diverges as one goes back in time. This means that different regions in the early universe could have communicated with each other, and come to equilibrium at a common state. So explaining why the microwaves, look the same in different directions.

The original model of inflation, which came to be known as old inflation, had various problems. How did the universe get into a false vacuum state in the first place, and how did it get out again? Various modifications were proposed, that went under the names of new inflation, or extended inflation. I won't describe them, because I have got into trouble in the past, about who should have credit for what, and because I now consider them irrelevant.

As Linde first pointed out, it is not necessary for the universe to be in a false vacuum, to get inflation. A scalar field with a potential V, will have the energy momentum tensor shown below.

$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2} g_{\mu\nu}[\phi_{,\Lambda}\phi^{,\Lambda} + V(\phi)] \tag{2}$$

Energy momentum tensor.

If the field is nearly constant in a region, the gradient terms will be small, and the energy momentum tensor, will be minus half V, times the metric. This is just what one needs for inflation.

In the false vacuum case, the scalar field sits in a local minimum of the potential, V. In that case, the field equation allows the scalar field, to remain constant in space and time. If the scalar field is not at a local minimum, it can not remain constant in time, even if it is initially constant in space. However, Linde pointed out that if the potential is not too steep, the expansion of the universe, will slow down the rate at which the field rolls down the potential, to the minimum. The gradient terms in the energy momentum tensor, will remain small, and the scale factor will increase almost exponentially. One can get inflation with any reasonable potential V, even if it didn't have local minima, corresponding to false vacua. The work that Neil and I have done, is a logical extension of Andrei's idea. But I'm not sure if Andrei agrees with it, though I think he's coming round.

Andrei's idea removed the need to believe that the universe began in a false vacuum. However, one still needed to explain, why the field should have been nearly constant over a region, with a value that was not at the minimum of the potential. To do this, one has to have a theory of the initial conditions of the universe. There are three main candidates.

They are, the so called pre-big bang scenario, the tunneling hypothesis, and the no boundary proposal. In my opinion, the pre-big bang scenario is misguided, and without predictive power. And I feel the tunneling hypothesis, is either not well defined, or gives the wrong answers. But then I'm biased, for it was Jim Hartle and I, that were responsible for the no boundary proposal.

This says that the quantum state of the universe, is defined by a Euclidean path integral over compact metrics, without boundary.

One can picture these metrics, as being like the surface of the Earth, with degrees of latitude, playing the role of imaginary time. One starts at the north pole, with the universe as a single point. As one goes south, the spatial size of the universe, increases like the lengths of the circles of latitude. The spatial size of the universe, reaches a maximum size at the equator, and then shrinks again to a point at the south pole.

Of course, spacetime is four dimensional, not two dimensional, like the surface of the Earth, but the idea is much the same. I shall go through it in detail, because it is basic to the work I'm going to describe.

The simplest compact four dimensional metric that might represent the universe, is the four sphere.

One can give its metric in terms of a coordinate, σ, that measures the distance from the pole, and three coordinates, χ, θ and ϕ, on a three sphere, that represents the spatial size of the universe. Again, one starts at the north pole, $\sigma = 0$, with a universe of zero spatial size, and expands up to a maximum size at the equator, $\sigma = \pi$, over 2H. But we live in a universe with a Lorentzian metric, like Minkowski space, not a Euclidean, positive definite metric. One therefore has to analytically continue, the Euclidean metrics used in the path integral, for the no boundary proposal. There are several ways one can analytically continue, the metric of the four sphere, to a Lorentzian spacetime metric.

$$ds^2 = d\sigma^2 + \frac{1}{H^2} \sin^2 H\sigma \, (d\chi^2 + \sin^2\chi \, d\Omega^2)$$

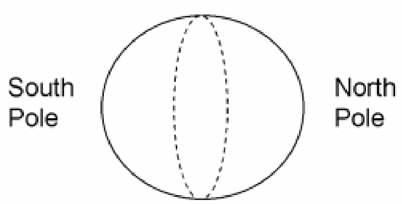


Fig. 2. The simplest compact four dimensional metric that might represent the universe; the four sphere.

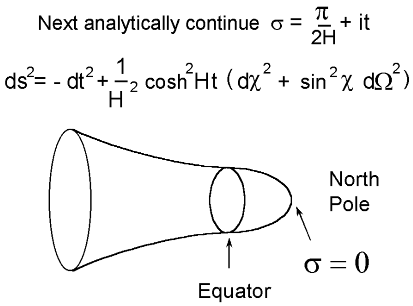


Fig. 3.

The most obvious is to follow the Euclidean time variable, σ, from the north pole to the equator, and then go in the imaginary σ direction, and call that real Lorentzian time, t.

Instead of the size of the three spheres going as the $\sin(H\sigma)$, they now go as the $\cosh(Ht)$. This gives a closed universe, that expands exponentially with real time. At late times, the expansion will change from being exponential, to being slowed down by matter in the normal way. This departure of the scale factor from a cosh behavior, will occur because the original Euclidean four sphere, was not perfectly round. But the universe would still be closed, however deformed the four sphere.

For nearly 15 years, I believed that the no boundary proposal, predicted that the universe was spatially closed.

$$\Omega_{matter} + \Omega_\Lambda = 1 + \tfrac{\kappa}{3\dot{s}^2} \tag{3}$$

Einstein equations.

But the Einstein equations, relate the energy density in the universe, plus lambda, to the rate of expansion, and the curvature, k, of the surfaces of constant time. Define Ω_{matter} and Ω^Λ, to be the density and Λ, divided by the critical value. If the universe is closed, that is, k=+1, Ω_{matter} plus Ω^Λ, must be greater than one. Observations seemed to show that omega matter at least, was significantly less than one. Still, Eddington once said, if your theory doesn't agree with the observations, don't worry. The observations are probably wrong. But if your theory doesn't agree with the second law of thermodynamics, forget it. I firmly believed in the no boundary proposal, and I thought it implied that the universe had to be closed. Since a closed universe, is not incompatible with the second law of thermodynamics, I was sure the observers had missed something, and there really was enough matter to close the universe. At that time, I didn't take seriously the possibility of a small cosmological constant.

You will hear from other people at this conference, about observations of Ω. In my opinion, they do not yet show that the universe is definitely open,

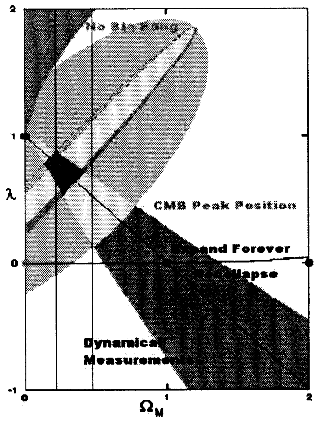


Fig. 4. Summary of observations.

or that Λ is non zero, but it is beginning to look like one or the other, if not both. One can summarize the observations in this diagram.

The vertical strip, corresponds to measurements of gravitational clustering, which suggests that Ω_{matter}, is in the range of point 2, to point 4. The yellow, red and green areas, represent the formal errors of the supernova observations, and the large pink area, other possible errors. Also shown in blue, are the limits set by the position of a peak in the angular spectrum, of the variations of the microwave background. As you can see, the observations suggest that the universe is closed to the open close divide, but with a non zero Λ.

Despite these indications of a low density Λ universe, I continued to believe that the cosmological constant was zero, and the no boundary proposal, implied that the universe must be closed. Then in conversations with Neil Turok, I realized there was another way of looking at the no boundary universe, that made it appear open.

One starts with the point that Andrei Linde made, that inflation doesn't need a false vacuum, a local minimum of the potential. But if the scalar field is not at a stationary point of the potential, then it can not be constant on an instanton, a Euclidean solution of the field equations. In turn, this implies that the instanton can't be a perfectly round four sphere. A perfectly round four sphere, would have the symmetry group, O(5). But with a non constant scalar field, the largest symmetry group that an instanton can have, is O(4). In other words, the instanton is a deformed four sphere.

$$ds^2 = d\sigma^2 + b^2(\sigma)(d\chi^2 + \sin^2 \chi d\Omega^2) \tag{4}$$

O(4) Instanton.

One can write the metric of an O(4) instanton, in terms of a function, b of σ. Here b is the radius of a three sphere of constant distance, σ, from the north pole of the instanton. If the instanton were a perfectly round four sphere, b would be a sin function of σ. It would have one zero at the north pole, and a second at the south pole, which would also be a regular point of the geometry. However, if the scalar field at the north pole, is not at a stationary point of the potential, it will be almost constant over most of the four sphere, but will diverge near the south pole. This behavior is independent of he precise shape of the potential. The non constant scalar field, will cause the instanton not to be a perfectly round four sphere, and in fact there will be a singularity at the south pole. But it will be a very mild singularity, and the Euclidean action of the instanton will be finite.

This Euclidean instanton, has been described as the universe begining as a pea. In fact, a pea is quite a good image for a deformed sphere. Its size of a few thousand Planck lengths, makes it a very petty pwa. But the mass of the matter it contains, is about half a gram, which is about right for a pea.

I actually discovered this pea instanton in 1983, but I thought it could describe the birth of closed universes only.

$$ds^2 = -dt^2 + \frac{1}{H^2}\cosh^2 Ht(d\chi^2 + \sin^2\chi d\Omega^2) \tag{5}$$

Closed universe from the P-instanton.

To get a closed universe, one starts with $\sigma = 0$ at the north pole, and proceeds to the equator, or rather the value of σ at which the radius, b, of the three sphere is maximum. One then analytically continues sigma in the imaginary direction, as Lorentzian time. As I described earlier, this gives a closed universe with a scale factor that initially goes like $\cosh(t)$. The scalar field, will have a small imaginary part, but that can be corrected by giving the initial value of the scalar field at the north pole, a small imaginary part.

According to the no boundary proposal, the relative probability of such a closed universe, is e to minus twice the action of the part of the pea instanton, between the north pole, and the equator. Notice that as this part, doesn't contain the singularity at the south pole, there is no ambiguity about the action of a singular metric. The action of this part of the instanton, is negative, and is more negative, the bigger the pea. Thus the probability of the pea, is bigger, the bigger the pea. The negative sign of the action, may look counter intuitive, but it leads to physically reasonable consequences.

As I said, I thought the no boundary proposal, implied that the universe had to be spatially closed, and finite in size. But Neil Turok and I, realized his ideas on open inflation, could be fitted in with the no boundary proposal. The universe would still be closed and finite, in one way of looking at it. But in another, it would appear open and infinite.

$$ds^2 = -dt^2 + a^2(d\psi^2 + \sinh^2\psi d\Omega^2) \tag{6}$$

Open universe from the P-instanton.

Let's go back to the metric for the pea instanton, and analytically continue it in a different way. As before, one analytically continues the Euclidean latitude coordinate, in the imaginary direction, to become a Lorentzian time, t. The difference is that one goes in the imaginary sigma direction at the north pole, rather than the equator. One also continues the coordinate, χ, in the imaginary direction, as a coordinate, ψ. This changes the three sphere, into a hyperbolic space. One therefore gets an exponentially expanding open universe.

One can think of this open universe, as a bubble in a closed, de Sitter like universe. In this way, it is similar to the single bubble inflationary universes, that have been proposed by a number of authors.

The difference is, the previous models all required carefully adjusted potentials, with false vacuum local minima. But the pea instanton, will work for any reasonable potential. The price one pays for a general potential, is a singularity at the south pole. In the analytically continued Lorentzian space-time, this singularity would be time like, and naked.

One might think that this naked singularity, would mean one couldn't evaluate the action of the instanton, or of perturbations about it. This would mean that one couldn't predict the quantum fluctuations, or what would happen in the universe. However, the singularity at the south pole, the stalk of the pea, is so mild, that the actions of the instanton, and of perturbations around it, are well defined. At first, it seemed just a lucky accident, that the singularity had these properties, but it now appears that there may be a deep reason, that I will come on to later. This behavior of the singularity, means one can determine the relative probabilities of the instanton, and of perturbations around it. The action of the instanton itself, is negative, but

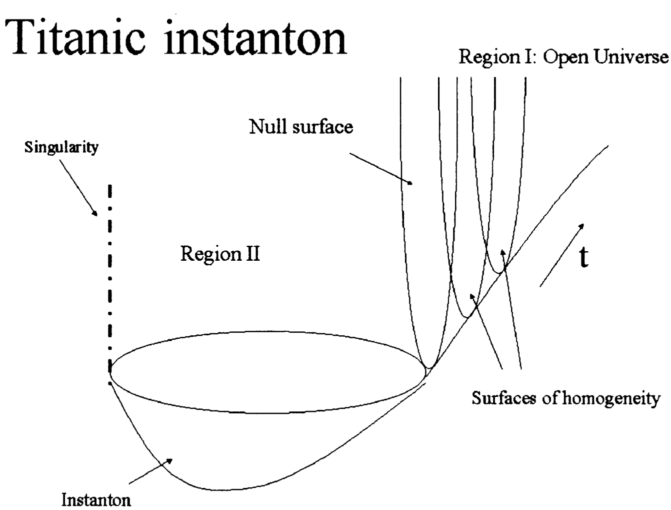


Fig. 5. The Titanic Instanton.

the effect of perturbations around the instanton, is to increase the action, that is, to make the action less negative.

$$Probability = e^{-2S_E} \tag{7}$$

No boundary proposal (Hawking Hartle)

$$Probability = e^{+2S_E} \tag{8}$$

Tunnelling hypothesis (Linde Vilenkin)

According to the no boundary proposal, the probability of a field configuration, is e to minus its action. Thus perturbations around the instanton, have a lower probability, than the unperturbed background. This means that quantum fluctuation are suppressed, the bigger the fluctuation, as one would hope. On the other hand, according to the tunneling hypothesis, favored by Vilenkin and Lindeh, probabilities are proportional to e to the plus action. This would mean that quantum fluctuation would be enhanced, the bigger the fluctuation. There is no way this could lead to a sensible description of the universe. Lindeh therefore proposes to take e to the plus action, for the probability of the background universe, but e to the minus action, for the perturbations. However, there is no invariant way, in which one can divide the action, into a background part, and a part due to fluctuations. So Linde's proposal, does not seem well defined in general. By contrast, the no boundary proposal, is well defined. Its predictions may be surprising, but they are not obviously wrong.

So we see that a general potential, without false vacuums, or local minima, leads to the P-instanton. This can be analytically continued, to either an open, or a closed universe. The no boundary proposal, then allows one to calculate the relative probabilities of different backgrounds, and the quantum fluctuations about them.

There isn't just a single P-instanton, but a whole family of them, labelled by different values of the scalar field at the north pole. The higher the value of the potential at the north pole, the smaller the instanton, and the less negative the value of the action. Thus the no boundary proposal, predicts that large instantons, are more probable than small ones. This is a problem, because large instantons, will lead to a shorter period of exponential expansion or inflation, than small ones. In the closed universe case, a short period of inflation, would mean the universe would recollapse before it reached the present size and density. On the other hand, an open universe with a short period of inflation, would become almost empty early on.

Clearly, the universe we live in, didn't collapse early on, or become almost empty. So we have to take account of the anthropic principle, that if the universe hadn't been suitable for our existence, we wouldn't be asking why it is, the way it is. Many physicists don't like the anthropic principle, but I think some version of it is essential, in quantum cosmology. M theory, or whatever the ultimate theory is, seems to allow a very large number of possible solutions, and compactifications. One has to have some criterion, for discarding most of them. Otherwise, why isn't the universe, eleven dimensional Minkowski space. According to the no boundary proposal, there should be an amplitude for any kind of spacetime, but most of them won't be of much interest, because they won't contain life.

The approach Neil Turok and I took, was to invoke the weakest version of the anthropic principle. We adopted Bayes statistics.

$$P(\Omega_{matter}, \Omega_{\Lambda}|Galaxy) \text{ is proportional to } P(Galaxy|\Omega_{matter}, \Omega_{\Lambda}) \times P(\Omega_{matter}, \Omega_{\Lambda}) \tag{9}$$

Bayesian statistics

In this, one starts with an a-priori probability distribution, and then modifies it in light of ones knowledge of the system. In this case, we took the a-priori distribution, to be the e to the minus action, predicted by the no boundary proposal. We then modified it, by the probability that the model contained galaxies, which are presumably a necessary condition, for the existence of intelligent observers. An open universe, has an infinite spatial volume. Thus the total number of galaxies in an open universe, would always be infinite, no matter how low the probability of finding a galaxy, in a given comoving volume. One therefore can not weight the a-priori probability, given by the no boundary proposal, by the total number of galaxies in the universe. Instead, we weighted by the comoving density of galaxies, predicted from the growth of quantum fluctuations, about the pea instanton. This gives a modified probability distribution for omega, the present density, divided by the critical density. For the open models, this probability distribution, is sharply peaked at an omega of about zero point zero one. This is lower than is compatible with the observations, but it is not such a bad miss. As far as I'm aware, this is the first attempt to predict a value of Ω for an open universe, rather than fine tune a false vacuum potential, to obtain a value in the range indicated by observation.

The reason we obtained such a low value for Ω, was that the a-priori probability distribution, given by the no boundary proposal, depended so steeply on omega. This meant that the modified probability distribution, was sharply peaked at the minimum value of omega, that would allow a single galaxy in our Hubble volume. Clearly this is not the universe we live in, which is just as well, because we cosmologists would all be out of a job, if there were no other no other galaxy visible. Refining the anthropic arguments, will not solve this problem. We just don't need all those other galaxies, for our own existence. The only way I can see of explaining omega without fine tuning, is if there is not a continuous family of instantons, with different values of omega, but only a discrete set. If the spacing in omega is large, there may be only one value of omega in the allowed range. Other values of Ω might have much higher a-priori probabilities, but they would correspond to uninhabited universes.

It seems there may be a way of geting such a discrete set, from eleven dimensional super gravity, which is the best candidate we have, for a theory of everything. It has a three form potential, with a four form field strength. When dimensionally reduced to four dimensions, this can act as a cosmological constant. For a real four form in four dimensions, the contribution to the cosmological constant is negative, which isn't what we want. However, super symmetry is broken, in the universe we live in.

Indeed, super symmetry breaking, is a necessary condition for life. Life could not develop in a super symmetric universe, filled with mass less particles. The breaking of super symmetry, will give rise to a positive contribution to the cosmological constant. This has been seen as a great problem, and there have been elaborate fine tuning schemes proposed, to get rid of it. But maybe the positive contribution from symmetry breaking, is canceled by the negative contribution from the four form. The fine tuning would be provided by the anthropic principle. Galaxies would not form, unless the total cosmological constant, was almost zero.

But symmetry breaking and the four form, need not cancel each other exactly. The anthropic requirement, can probably be satisfied by any omega lambda between about minus one point five, and plus one point five. This is consistent with the observations of omega lambda.

When I gave a talk on this in Santa Barbara, Joe Polchinski pointed out that the four form was not continuously variable, but that the integral of its dual over a seven cycle, was quantized. Unless the compact dimensions were unacceptably large, the spacing of the contribution to the cosmological constant, from different levels of the four form, was larger than the observational limit on lambda. Thus one could not adjust the four form, to cancel symmetry breaking, to the accuracy required. At the time, I saw this as a setback, but now I see it as a positive advantage. Coupled with another idea I shall describe, it could mean that there is not a continuous variation in the size of the instantons, with all the problems that causes with low omega, but only a discrete set.

The other idea I mentioned, is due to Garriga. He took a regular higher dimensional space, and dimensionally reduced it with respect to a U1 isometry group, that had a fixed point. He obtained a four dimensional geometry, with a scalar field that was the log of the length of the U1 fibers. Near the fixed point, the scalar field and four geometry behaved just like those in the pea instanton, near the south pole. So the apparent singularity in the pea instanton, may be resolved by going to higher dimensions. This could be the deep reason, why the singularity is so mild, that the solution, and perturbations around it, have well defined actions, so the quantum fluctuations, are well defined. Having a discrete set of non singular instantons, would be more in keeping with the spirit of the no boundary proposal. As I said earlier, it might also offer a way out of the problem of low omega. However, a lot more work needs to be done, to get a concrete model. The aim is to find a description of the origin of the universe, on the basis of fundamental theory.

Assuming that one can find a model that predicts a reasonable omega, how can we test it by observation. The best way is by observing the spectrum of fluctuations, in the microwave background. This is a very clean measurement of the quantum fluctuations, about the initial instanton. However, there is an important difference between our instanton, and previous proposals for open inflation. They have all assumed false vacuum potentials, and have used the Coleman De Lucia instanton, which is non singular. On the other hand, our instanton has a singularity at the south pole. As I said, quantum fluctuations around the instanton are well defined, despite the singularity. Perturbations of the Euclidean instanton, have finite action if and only, they obey a Dirichelet boundary condition at the singularity. Perturbation modes that don't obey this boundary condition, will have infinite action, and will be suppressed. The Dirichelet boundary condition also arises, if the singularity is resolved in higher dimensions.

When one analytically continues to Lorentzian spacetime, the Dirichelet boundary condition, implies that perturbations reflect at the time like singularity. This has an effect on the two point correlation function of the perturbations, but it seems to be quite small. The present observations of the microwave fluctuations, are certainly not sensitive enough to detect this effect. But it may be possible with the new observations that will be coming in, from the map satellite in two thousand and one, and the Planck satellite in two thousand and six. Thus the no boundary proposal, and the pea instanton, are real science. They can be falsified by observation.

I will finish on that note.

Generating Cosmological Solutions from Known Solutions

Hernando Quevedo and Michael P. Ryan, Jr.

Instituto de Ciencias Nucleares, UNAM, A. Postal 70-543,
México 04510 D.F., Mexico

Abstract. We consider three methods by which one can generate new cosmological models. Two of these are based on the Lorentzian structure of spacetime. In a Lorentzian manifold there can exist horizons that separate regions of spacetime that can be interpreted as cosmological models from others that have the character of "black holes." A number of well known solutions of this type can be used to generate both known cosmological models and others that do not seem to have been recognized. Another method based on the Lorentzian character of spacetime is to simply interchange some space variable with time and try to restructure the metric to make a viable cosmology.

A more broad-ranging method is the use of modern solution-generating techniques to construct new models. This method has been widely used to generate black hole solutions, but seems not to have been so widely used in cosmology. We will discuss examples of all three methods.

1 Introduction

The standard model of cosmology assumes that the universe today is homogeneous and isotropic, which means that there are six Killing vectors of the manifold that give the isometries that realize these symmetries. However, in earlier stages of the universe it is assumed that there might have been large amounts of anisotropy and inhomogeneity that, by some physical process, could have been reduced to the point where today we observe the standard model (for a compendium of reasons for considering inhomogeneous models, see Ref. [1]). Anisotropic and inhomogeneous metrics have fewer and fewer Killing vectors, and we could, in principle, arrive at a completely general solution of the Einstein equations as a cosmological model. At some point it will become difficult to distinguish between a cosmological model and any other general solution of the Einstein equations, except as a matter of interpretation. In fact, even an inhomogeneous cosmology with two spacelike Killing vectors is difficult to distinguish (locally) from gravitational waves propagating in one space direction. One method that has been used to separate cosmological models from pure wave solutions is the prescription of their global topology. If we insist that cosmologies have compact $t =$ constant surfaces, then a number of gravitational wave solutions can be made into cosmologies by compactifying in certain directions of the system. We will give

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 191–213, 2000.

examples of this procedure below. If we allow open universes, whether we call a solution a cosmological model or not is again a matter of interpretation.

With these caveats we can give a number of methods by means of which one can transform known solutions of the Einstein equations into cosmological models. These methods can be broken down into three broad classes:

1) Horizon methods
2) Causal structure methods
3) Mapping methods

The first of these methods has a long history, even though it never seems to have been thought of as a "method." Perhaps the oldest example is the deSitter metric [2], although it is a somewhat degenerate example. Originally deSitter proposed a solution to the Einstein equations with cosmological constant Λ of the form [3]

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

This seems to be a static metric similar in form to the Schwarzschild metric, and even has a "singularity" (which caused much comment at the time) at $r = \sqrt{3/\Lambda}$. Of course, this singularity is just a horizon where the light cones tip over sufficiently that for $r > \sqrt{3/\Lambda}$, $\partial/\partial r$ becomes timelike and $\partial/\partial t$ spacelike. For large r, then, we can rename the coordinates ($r \to \tilde{t}$, $t \to \tilde{r}$), and the new metric is

$$ds^2 = -\frac{d\tilde{t}^2}{\frac{\Lambda\tilde{t}^2}{3} - 1} + \left(\frac{\Lambda\tilde{t}^2}{3} - 1\right) d\tilde{r}^2 + \tilde{t}^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{2}$$

which is an obvious cosmological model. It was "recognized" (the reason for using the quotation marks will become obvious below) relatively quickly that this was indeed a cosmological model by making use of the coordinate transformation

$$\begin{gathered}\tilde{t} = re^{\sqrt{\Lambda/3}T}, \quad \tilde{r} = T - \frac{1}{2}\sqrt{\frac{3}{\Lambda}} \ln\left(\frac{\Lambda r^2 e^{2\sqrt{\Lambda/3}T}}{3} - 1\right), \\ r\sin\theta\cos\varphi = X, \qquad r\sin\theta\sin\varphi = Y, \qquad r\cos\theta = Z\,,\end{gathered} \tag{3}$$

which makes the metric

$$ds^2 = -dT^2 + e^{2\sqrt{\Lambda/3}T}[dX^2 + dY^2 + dZ^2], \tag{4}$$

which is a $k = 0$ Friedmann-Robertson-Walker metric.

Here we should study the Penrose diagram of this metric. In Figure 1 we give a Penrose diagram that will cover several cases that we will discuss later. Because of this we will give only generic values in the figure, and each

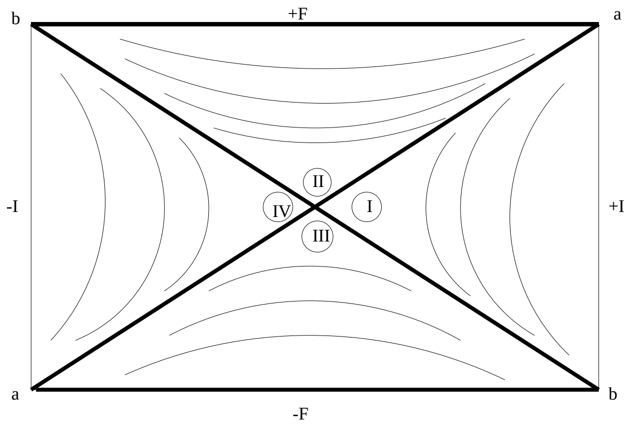


Fig. 1. A generic Penrose diagram which will be used for several metrics of the article

case will correspond to different values of these parameters. In the deSitter case, the lines aa and bb correspond to $r = \sqrt{3/\Lambda}$, and both $\pm I$ to $r = 0$. In region I the spacelike Killing vector $\partial/\partial r$ generates spacelike surfaces that are the t = constant surfaces of (1). In region II the same Killing vector is now $\partial/\partial\tilde{t}$ (timelike) and generates timelike hypersurfaces $\tilde{r}$ = constant. The horizon splits the region comprised by both regions I and II in two. Region I might be called a "black hole" region since the metric is reminiscent of the Schwarzschild metric, and has the spacelike Killing vector $\partial/\partial r$, while region II is a "cosmological" region with metric (4). In the cosmological region there seem to be "singularities" at $T = \pm\infty$, but we can see that [note that $T = \tilde{r} + \frac{1}{2}\sqrt{\frac{3}{\Lambda}} \ln\left(\frac{\Lambda\tilde{t}^2}{3} - 1\right)$] the singularity at $T = -\infty$ is just the horizon at $\tilde{t} = \sqrt{3/\Lambda}$, so it is nothing more than a null surface separating two parts of the spacetime.

The problem here is that the deSitter metric has such a high degree of symmetry (a space of constant curvature with 10 Killing vectors) that the distinction between regions I and II is artificial. In fact, the coordinate transformation (3) (with $\tilde{t} \to r$ and $\tilde{r} \to t$ and $\frac{\Lambda\tilde{t}^2}{3} - 1 \to \frac{1-\Lambda r^2}{3}$) is equally valid in region I, and, in fact, this was what was recognized by Lemaitre [4] and Robertson [5] in transforming the deSitter metric into the form that was eventually used in the steady-state model. Actually, even Minkowski space (also a space of constant curvature with 10 Killing vectors) can be looked at in the same way, as we will see in Sect. 2.

In Sect. 2 we will consider several metrics where horizon methods can be used to construct cosmological models that do not have the disadvantages of the deSitter metric, that is, there exists no coordinate transformation that is valid for the static region which transforms it into a cosmological model. The fact that "singularities" in the new cosmological models may just be null surfaces where one passes from one coordinate patch to another, will, of course, still be a feature, although true curvature singularities may also exist. One point that should be mentioned is that most of the models generated by this method that we will consider are vacuum models. There are some electrovac models that can be thought of as being generated by horizon methods that we will mention, but it seems not to be known which matter-filled cosmological models can be generated by this method.

The second class of methods, causal structure methods, is similar to the previous one, but there is no horizon to distinguish different regions of spacetime. The paradigm of this method is the unpolarized Gowdy T^3 vacuum spacetime [6] which will be discussed in more detail below. The idea behind this method is to begin with a known solution and simply rename some spacelike variable as t, and the timelike variable as a spacelike variable (we can call this $t \leftrightarrow r$), and then change the causal structure of the resulting metric so that the a vector tangent to the "time" direction is truly timelike, and one tangent to the new "space" direction is spacelike. There are a number of problems with this approach. The first is that determining the causal structure change is not necessarily trivial. In a diagonal metric this difficulty can usually be handled by simply changing the signs of two of the metric components, but in a more general metric the procedure may be more complicated. The most important difficulty, however, is that the new metric is not guaranteed to be a solution to the Einstein equations. For matter filled models, where physical quantities can be used to define spacelike and timelike surfaces, the new metric will almost certainly not be a solution. For vacuum metrics it is difficult to tell. As with horizon methods, the known solutions of this type are vacuum wave solutions, and they have rather simple structures, and give inhomogeneous cosmological models. Here, of course, one has the problem of distinguishing between gravitational wave solutions and inhomogeneous models.

A very simple example of this procedure is the plane wave [7]

$$ds^2 = L^2(u)(e^{2\beta(u)}dx^2 + e^{-2\beta(u)}dy^2) + dz^2 - dt^2, \tag{5}$$

where $u = t - z$. If we make the coordinate change $t \to \tilde{z}$, $z \to \tilde{t}$ and change the signs of dz^2 and dt^2, we find ($\tilde{u} = \tilde{t} - \tilde{z}$)

$$ds^2 = L^2(-\tilde{u})[e^{2\beta(-\tilde{u})}dx^2 + e^{-2\beta(-\tilde{u})}dy^2] - d\tilde{t}^2 + d\tilde{z}^2. \tag{6}$$

Since L and β are arbitrary function of their arguments, we can remove the minus signs in the arguments of L and β in (6) and obtain (5) again. Because of the high symmetry of the problem, this procedure gives the same metric and Einstein equations, and there is no question that any solution of the Einstein equations for (5) gives a solution to (6). Here, of course, one runs into the problem of defining a cosmological model. Is either of (5) or (6) a gravitational wave or an inhomogeneous cosmological model? Either one is both, depending on how we interpret them. If, however, we insist that a cosmological model have compact $t = \text{const.}$ sections, then we can compactify (6) in the space directions by just making $0 \leq x, y, z \leq 2\pi$ and identifying the points at 0 and 2π, giving the manifold a T^3 topology. This changes the boundary conditions on the functions L and β in that one must have $L(t, 0) = L(t, 2\pi)$ and $\beta(t, 0) = \beta(t, 2\pi)$. Since the only Einstein equation is

$$L''(\tilde{u}) + [\beta'(\tilde{u})^2]L = 0, \tag{7}$$

$' = d/d\tilde{u}$, β is arbitrary except for the boundary conditions, and L obeys a simple equation that has solutions that satisfy the boundary conditions. Here one must be careful not to change coordinates back to the original metric, since we will generate a spacetime with closed timelike lines, a problem with cosmological models generated in this manner. We will discuss examples and their topology in Sect. 2.

The third method consists of taking known solutions and mapping them to new solutions. This technique is similar in spirit to conformal mapping. Any analytic function of one complex variable, $w = f(z)$ maps $z = x + iy$ to $w = u(x, y) + iv(x, y)$, where, if $\psi(x, y)$ is a solution to Laplace's equation, $\Psi[u(x, y), v(x, y)]$ is also a solution. This idea exploits the invariance of Laplace's equation in two dimensions under the two-dimensional conformal group. If one can find groups under which the Einstein equations are invariant (or under which some class of Einstein equations for special systems are invariant), one can generate new solutions from old. This technique is of wide applicability, and has been exploited heavily in general relativity to find new exact solutions. It is by no means specific to cosmological models, but it can be applied in cosmological situations, and has, perhaps, been underutilized in this field and deserves more attention there.

In the specific case of stationary axisymmetric spacetimes, early generating methods such as the Kerr-Schild *ansatz* [8,9,10,11], the complex transformation method [12,13,14], and Hamilton-Jacobi separability [15] were used to derive new solutions, but all of them are contained in the charged Kerr-Taub-NUT [16] class (with cosmological constant). An important development for the derivation of exact stationary axisymmetric solutions was made by Ernst [17,18], who obtained a new representation of the corresponding field equations which is independent of the coordinates chosen and, therefore, allows one to investigate the symmetries involved and find new solutions. Ernst also proposed two generating techniques that were later enlarged and unified by Kinnersley [19,20]. Using the compact Ernst formulation of the field equations it was also possible to obtain new solutions [21,22,23,24, 25] possessing a certain prescribed polynomial form of the Ernst potential. Modern solution generating techniques involve Lie groups of transformations or Bäcklund transformations. The first such group was found by Geroch [26, 27] and generalized by Cosgrove [28,29]. The Geroch group was investigated very intensively and it was found that it possesses subgroups that preserve asymptotic flatness [19,30,31,32,33]. Today, there exist three main solution generating techniques: HKX (Hoenselaers, Kinnersley, Xanthopoulos) transformations [34,35], which are based upon special subgroups of the Geroch group, Bäcklund transformations [36,37], which are applied directly to the Ernst potential, and the inverse scattering method [38,39], which is based on a reformulation of the nonlinear field equations as a linear eigenvalue problem. Later on it was shown [40,41] that all these techniques are related to one another, and can be used to generate the same type of solutions (for an introductory review of the solution generating techniques and the relationships between them see [42]).

The mapping method is by no means specific to stationary axisymmetric metrics, but it can be applied to any spacetime characterized by two or more commuting Killing vector fields as in the case of the cosmological models investigated in this work. We will give some examples in Sect. 3 and a guide to possible new uses of the technique.

The paper is organized as follows. Section 2 will give examples of horizon and causal structure methods and solutions obtained using them. Section 3 will give one example of mapping methods and discuss others. Finally, Sect. 4 will discuss possible new directions in the use of known solutions to find new cosmologies.

2 Horizon and Causal Structure Methods

2.1 Horizon Methods

There are a large number of solutions of the Einstein equations that have cosmological coordinate patches as well as "black hole" regions that are separated by horizons. The major problem with the cosmologies generated in this way is they are incomplete manifolds, and what have in the past been interpreted in some of them as singularities are just the horizons that separate one part of the complete manifold from the other. A second problem, as we have seen, is that in certain cases the complete manifold is of such high symmetry that is is impossible to distinguish between the cosmological region and the "black hole" region. The most blatant example of the second problem is just ordinary Minkowski space.

If we return to Figure 1, we can think of this diagram as a picture of Minkowski space with the meeting point of lines aa and bb an arbitrarily chosen origin O. The usual metric is

$$ds^2 = -dT^2 + dx^2 + dy^2 + dz^2. \tag{8}$$

In region I the hyperbolic lines are just the usual orbits of a particle with constant acceleration that form the basis of Rindler [43] space, and if the lines are given by $x = \sqrt{2r-1}\cosh t$ and $T = \sqrt{2r-1}\sinh t$, $z = y = \text{const.}$, the metric becomes

$$ds^2 = -(2r-1)dt^2 + \frac{dr^2}{2r-1} + dy^2 + dz^2. \tag{9}$$

In this coordinate system the metric is very similar in form to the Schwarzschild metric, and there is a coordinate singularity at $r = 1/2$, which is a horizon similar to that of the deSitter metric.

In Figure 1 the line bb is the surface $r = 1/2$, and for $r < 1/2$ we can make the $t \leftrightarrow r$ coordinate change and the metric becomes

$$ds^2 = -\frac{1}{1-2t}dt^2 + (1-2t)dr^2 + dy^2 + dz^2, \tag{10}$$

a cosmological model. If we change coordinates using $t = \frac{1}{2}[1 - \tau^2\{\cosh^2\rho - \sinh^2\rho\sin^2\theta\cos^2\varphi\}]$, $r = \tanh^{-1}[\tanh\rho\sin\theta\cos\varphi]$, $y = \tau\sinh\rho\sin\theta\sin\varphi$, $z = \tau\sinh\rho\cos\theta$, the metric (10) becomes the Milne universe, a flat cosmological model with metric

$$ds^2 = -d\tau^2 + \tau^2[d\rho^2 + \sinh^2\rho(d\theta^2 + \sin^2\theta d\varphi^2)], \tag{11}$$

a $k = -1$ FRW metric. Of course, the manifold is nothing more than flat space and the breaking up of the manifold into regions I-IV is an observer-dependent phenomenon due to singling out the origin O as the position of a special observer, while there is actually no physical reason that this point is priveleged over any other point of the manifold, and there is no real difference between any of the four regions. Basically, we can say that the same is true of deSitter space, which is just a constant curvature space everywhere, and the “cosmological” region has nothing to distinguish it from any other region of the manifold.

There are a number of metrics which are not as degenerate as the Minkowski and deSitter cases whose cosmological regions are physically distinguishable from the “black hole” regions, and the resulting cosmological models have long been known and are named. Perhaps the simplest of these is the Kantowski-Sachs-Schwarzschild manifold. Here we can still use the generic Figure 1 to represent the Penrose diagram of this metric, with region I representing the Schwarzschild coordinate patch where the hyperbolic lines are $r =$ const. lines and the lines bb is the horizon at $r = 2m$. In region II the hyperbolic lines are still $r =$ const. lines, but they no longer represent spacelike surfaces. For $r > 2m$ the metric can be written as

$$ds^2 = -\left(1 - \frac{2m}{r}\right)dt^2 + \frac{1}{1-\frac{2m}{r}}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{12}$$

but for $r < 2m$ the Killing vector $\partial/\partial r$ is no longer spacelike, and we may make the transformation $r \leftrightarrow t$ mentioned in Sect. 1, and we find

$$ds^2 = -\frac{1}{\frac{2m}{t}-1}dt^2 + \left(\frac{2m}{t} - 1\right)dr^2 + t^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{13}$$

which is an obvious cosmological model, the vacuum Kantowski-Sachs model. The fact that at $r = 0$ ($t = 0$ in the new coordinate system) which is represented by the line at $+F$ is a true curvature singularity distinguishes region II from region I and makes it impossible to find a global coordinate transformation that makes this cosmological model indistinguishable from the "black hole" region. Of course, however, this cosmological model, and all cosmological models generated by horizon methods, have the problem that what has been regarded as singularities in these models may not be a curvature singularity, but only a horizon where we pass from one coordinate patch to another. In the case of the Kantowski-Sachs solution the $t = 0$ singularity is one of the usual singularities of the model, and it is a true curvature singularity. The other "singularity" at $t = 2m$ is only a null surface.

While the Kantowski-Sachs solution, even though it was originally discovered by means of studies of groups of motion that were not transitive on three surfaces, and constituted a generalization of the Bianchi cosmological models, and was immediately recognized as the Schwarzschild solution inside the horizon, in the next model we will discuss the two regions on either side of the horizon were studied separately. The "black hole" region was the NUT space of Newman, Tamburino and Unti [9], and the cosmological region was the Taub cosmology [44]. Misner [45] showed that these two solutions could be seen to be part of a larger manifold (up to topological questions which will be discussed below). In the black hole region one can write the NUT metric as [46]

$$ds^2 = -\frac{r^2 - 2mr - l^2}{r^2 + l^2}(dt + 2l\cos\theta d\varphi)^2 + \frac{r^2 + l^2}{r^2 - 2mr - l^2}dr^2 \\ +(r^2 + l^2)(d\theta^2 + \sin^2\theta d\varphi^2). \tag{14}$$

For the region of this metric where $0 < r < m(1 + \sqrt{1 + l^2/m^2})$, the $r \leftrightarrow t$ transformation gives

$$ds^2 = \frac{l^2 + 2mt - t^2}{t^2 + l^2}(dr + 2l\cos\theta d\varphi)^2 - \frac{t^2 + l^2}{l^2 + 2mt - t^2}dt^2 \\ +(t^2 + l^2)(d\theta^2 + \sin^2\theta d\varphi^2), \tag{15}$$

which is an anisotropic cosmological model. The major problem with this model is that in the cosmological sector we can rewrite the metric on $t =$ const. surfaces by making the coordinate transformation

$$y = \theta - \pi/2\,, \tag{16a}$$

$$z = r/2l\,, \tag{16b}$$

$$x = \varphi\,, \tag{16c}$$

which gives the three-dimensional line element as

$$d\sigma^2 = \frac{4l^2(l^2 + 2mt - t^2)}{t^2 + l^2}(dz - \sin y dx)^2 + (t^2 + l^2)(dy^2 + \cos^2 y dx^2), \tag{17}$$

which can be written as

$$d\sigma^2 = \frac{4l^2(l^2 + 2mt - t^2)}{t^2 + l^2}(\omega^3)^2 + (t^2 + l^2)[(\omega^1)^2 + (\omega^2)^2], \tag{18}$$

where $\omega^3 = dz - \sin y dx$, $\omega^1 = \cos y \cos z dx - \sin z dy$, $\omega^2 = \cos y \sin z dx + \cos z dy$. These differential forms are the invariant one-forms of the Bianchi type IX cosmological models, the invariant one-forms on S^3. This means that the "natural" topology of the three-surfaces given by (18) is that of S^3, with $0 \leq z \leq 4\pi$, $0 \leq x \leq 2\pi$, $0 \leq y \leq \pi$. If one maintains this topology on passing through the horizon, the vector $\partial/\partial z$ becomes timelike, and the S^3 topology implies the possibility of closed timelike lines. This topological difficulty is present in many cosmological models generated by horizon methods, as we will see below.

The metrics we have discussed seem to be the only ones where the cosmological models have been noticed explicitly to be a part of a larger manifold that has non-cosmological sectors. For example, it is only recently that the Kerr metric inside its horizon has been considered as a cosmological model. Of course, the metric in this part of the Kerr manifold has been studied [47], but there seems to have been no attempt to identify it as a cosmological model. In Boyer-Lindquist coordinates the Kerr metric has the form

$$\begin{aligned} ds^2 = &-\frac{r^2 - 2Mr + a^2}{r^2 + a^2\cos^2\theta}[dt - a\sin^2\theta d\phi]^2 + \frac{\sin^2\theta}{r^2 + a^2\cos^2\theta}[(r^2 + a^2)d\phi - adt]^2 \\ &+\frac{r^2 + a^2\cos^2\theta}{r^2 - 2Mr + a^2}dr^2 + (r^2 + a^2\cos^2\theta)d\theta^2. \end{aligned} \tag{19}$$

This metric has two horizons, $r_\pm = M \pm \sqrt{M^2 - a^2}$, and beyond the outer horizon, r_+, $\partial/\partial r$ is a spacelike vector, and the metric represents a spinning black hole. At r_+ the light cones tip over to the point where $\partial/\partial r$ becomes timelike and we can make the transformation $t \leftrightarrow r$. Unfortunately, at the inner horizon r_- the light cones tip back to the point where $\partial/\partial r$ becomes spacelike again. In the region $r_- < r < r_+$ the Kerr metric becomes a cosmological model,

$$\begin{aligned} ds^2 = &\frac{2Mt - t^2 - a^2}{t^2 + a^2\cos^2\theta}[dr - a\sin^2\theta d\phi]^2 + \frac{\sin^2\theta}{t^2 + a^2\cos^2\theta}[(t^2 + a^2)d\phi - adr]^2 \\ &-\frac{t^2 + a^2\cos^2\theta}{2Mt - t^2 - a^2}dt^2 + (t^2 + a^2\cos^2\theta)d\theta^2. \end{aligned} \tag{20}$$

This metric, with the simple transformation [48],

$$t = \alpha[\sqrt{1-\beta^2}\cos(e^{-\tau}) + 1], \tag{21}$$

$\alpha = M$, $\beta = a/M$ $(0 \leq \beta \leq 1)$ transforms (20) into

$$\begin{aligned} ds^2 = &e^{-\lambda/2}e^{\tau/2}(-e^{-2\tau}d\tau^2 + d\theta^2) + \alpha\sqrt{1-\beta^2}\sin(e^{-\tau})[e^P d\delta^2 \\ &+2e^P Q d\delta d\phi + (e^P Q^2 + e^{-P}\sin^2\theta)d\phi^2], \end{aligned} \tag{22}$$

where

$$\lambda = \tau - 2\ln(\alpha^2\{[\sqrt{1-\beta^2}\cos(e^{-\tau}) + 1]^2 + \beta^2\cos^2\theta\}), \tag{23}$$

$$P = \ln[(1-\beta^2)\sin^2(e^{-\tau}) + \beta^2\sin^2\theta] - \ln[\alpha\sqrt{1-\beta^2}\sin(e^{-\tau})]$$
$$- \ln([\sqrt{1-\beta^2}\cos(e^{-\tau}) + 1]^2 + \beta^2\cos^2\theta), \tag{24}$$

$$Q = -\frac{2\alpha\beta\sin^2\theta[\sqrt{1-\beta^2}\cos(e^{-\tau}) + 1]}{(1-\beta^2)\sin^2(e^{-\tau}) + \beta^2\sin^2\theta}\,. \tag{25}$$

The metric (22) is a Gowdy cosmological model with $S^1 \times S^2$ topology [6].

The solution given in Eqs. (23 -25) is a function of θ for $0 \leq \theta \leq \pi$ and is valid for values of τ that correspond to the part of the manifold between r_- and r_+, that is, for $-\ln(\pi) \leq \tau \leq +\infty$, (note that t takes the values $M \pm \sqrt{M^2 - a^2}$ at the two limiting values of τ).

As an example, Figs. 2 and 3 give $P-h$ ($h = -\ln[\alpha\sqrt{1-\beta^2}\sin(e^{-\tau})]$) and Q as functions of θ for various values of τ for $\alpha = 1$, $\beta = 1/2$. The evolution of these two functions is that of a "spike" that at the limiting values of τ is practically flat (except at $\theta = 0, \pi$, where it drops off drastically), and which becomes sharper for intermediate values of τ. It should be mentioned that the Kerr metric is a special case of what is usually called the Kerr-Taub-NUT metric, which is geven in Ref. [46]. This metric has three parameters, a, m, l, the Kerr parameter, the mass, and the NUT parameter respectively. For $a = 0$ we have the metric (14), and for $l = 0$ we have the Kerr metric. This means that the Gowdy model we have given is also a special case of the "Taub" region of Kerr-Taub-NUT. However, the Taub region of this manifold, in spite of its name, never seems to have been considered as a cosmological model.

As we have stated several times, the models studied so far are vacuum models. While matter-filled models may, in principle, be generated by horizon methods, in at least some cases the models with matter have very different behavior from the vacuum case, and may preclude naive generalizations of the method to the non-vacuum case. For instance, it is known that the Taub model with fluid matter has true curvature singularities where the vacuum case has null surfaces [49]. The same is true for the Kantowski-Sachs model [50]. However, the Brill model [51] is a Taub-NUT electrovac solution which is non-singular in the same way as the Taub-NUT vacuum case. In fact, the Reissner-Nordstrøm solution with electric charge large enough has two horizons, and between the outer and inner horizons there is a region that can be interpreted as an electrovac cosmology that has no curvature or electromagnetic field singularity at the horizons.

Besides the solutions we have mentioned, there are numerous metrics that represent gravitational fields outside some material body which in vacuum have "black hole" sectors outside of some horizon. Of course, no hair theorems [52] say that there are no true vacuum black holes that are not Kerr or Schwarzschild. Solutions of the type mentioned above seem always to have curvature singularities on or outside their horizons where they can be seen by observers at infinity. While this fact makes them poor candidates for black holes, it makes them interesting as cosmological models inside their

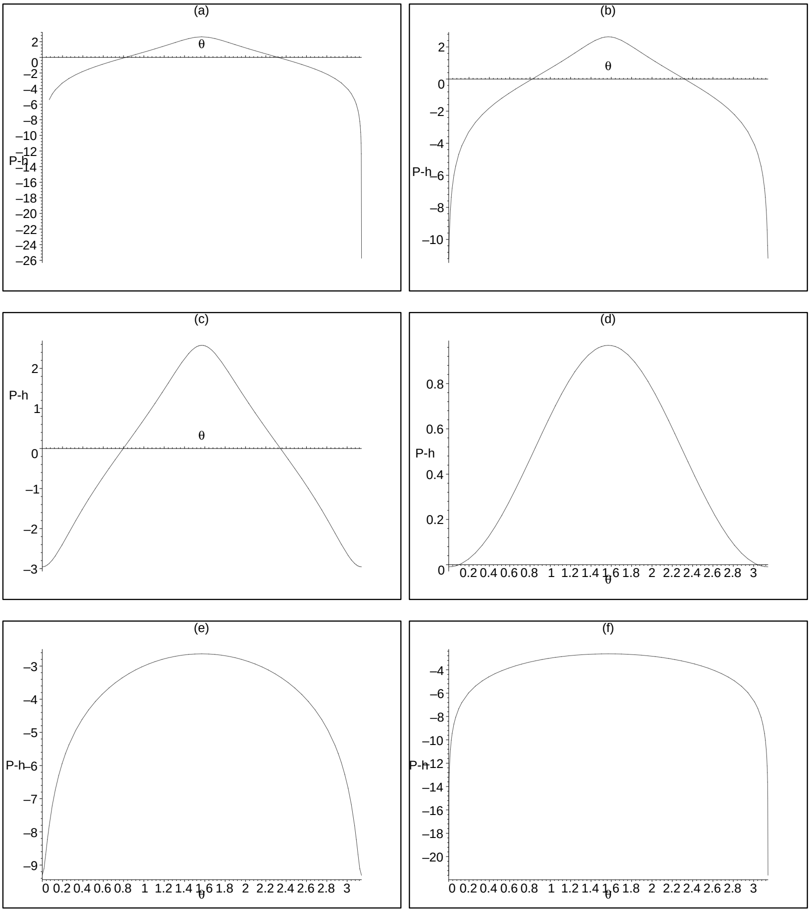


Fig. 2. The evolution in τ of $P - h$ as a function of θ from Eqs. (23-25) for $\alpha = 1$, $\beta = 1/2$. Fig. (a) corresponds to $\tau = -\ln(\pi)$, (b) to $\tau = -1.144$, (c) to $\tau = -1.1$, (d) to $\tau = -0.75$, (e) to $\tau = +5$, (f) to $\tau = +10$

horizons. The "singularities" of the models with curvature singularities on their horizons will be inhomogeneous, with part of the singularities being simple horizons, but other regions will be true curvature singularities. These structures might be of great interest in the study of cosmological singularities. Perhaps the simplest of the metrics of this type might be the Tomimatsu-Sato [21,22] models which do have inhomogenous horizons, but there are other known solutions, for example, one with many multipoles and large curvature singularity regions on the horizon [53].

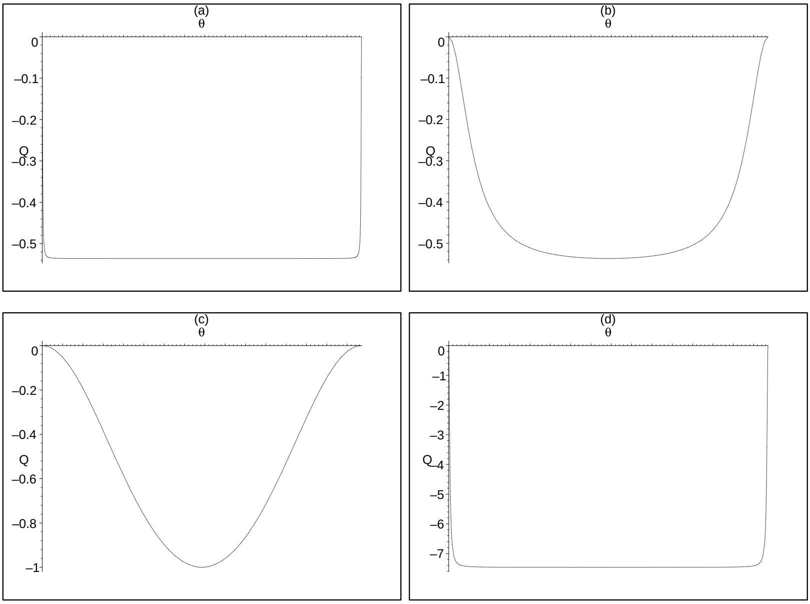


Fig. 3. The evolution of Q as a function of θ in τ. Fig. (a) corresponds to $\tau = -1.1439$, (b) to $\tau = -1.1$, (c) to $\tau = -\ln(\pi/2)$, (d) to $\tau = +5$

2.2 Causal Structure Methods

This method seems to have been little used, and perhaps the most telling reason for this is that there are very few examples where it works. The only example that has wide currency is that of certain Gowdy models. If one begins with the Einstein-Rosen waves [54], which have the metric

$$ds^2 = e^{2\gamma-2\psi}(dr^2 - dt^2) + r^2 e^{-2\psi} d\phi^2 + e^{2\psi} dz^2, \tag{26}$$

with γ and ψ functions of r and t, the Einstein equations for this metric are

$$\psi_{,rr} + \frac{1}{r}\psi_{,r} - \psi_{,tt} = 0, \tag{27}$$

$$\gamma_{,r} = r(\psi_{,r}^2 + \psi_{,t}^2), \qquad \gamma_{,t} = 2r\psi_{,r}\psi_{,t}, \tag{28}$$

where it is well known that the equations for γ can be integrated directly once ψ is known, since the integrability condition for these two first-order partial differential equations is just (27).

If we now simply make the coordinate change $t \leftrightarrow r$, we have

$$ds^2 = e^{2\gamma-2\psi}(dt^2 - dr^2) + t^2 e^{-2\psi} d\phi^2 + e^{2\psi} dz^2. \tag{29}$$

Here the only obstacle to interpreting this metric as a cosmological model is the wrong signs in the dr^2 and dt^2 terms if we want to interpret $\partial/\partial r$ and $\partial/\partial t$ as spacelike and timelike vectors respectively. If we add a complex constant to γ, $\gamma = \tilde{\gamma} + i\pi$, we change the signs of g_{tt} and g_{rr}, and since we have done nothing more than change the names of r and t, the Einstein equations are unchanged (since they depend only on derivatives of γ), and we have

$$-\psi_{,tt} - \frac{1}{t}\psi_{,t} + \psi_{,rr} = 0, \tag{30}$$

$$\tilde{\gamma}_{,t} = t(\psi_{,t}^2 + \psi_{,r}^2), \qquad \tilde{\gamma}_{,r} = 2t\psi_{,t}\psi_{,r} \tag{31}$$

These are the original equations for the unpolarized Gowdy T^3 cosmological model [6].

Notice that in order to change the sign of the $dt^2 - dr^2$ term, it was necessary to add $-i\pi$ to γ, and if the Einstein equations were to depend on γ in any other way than through derivatives or e^γ, they would become complex and then would not necessarily have real solutions for γ and ψ. This difficulty is a paradigm for the problems one would encounter in trying to use a $t \leftrightarrow r$ type of coordinate change. One would not expect this technique to be successful in the great majority of cases, especially if there were matter present. This seems to be reflected in the fact that there exist few examples of this method. Note that the examples we have given, the plane wave of the Introduction and the T^3 Gowdy model, are vacuum spacetimes, where the temporal variable and one of the space variables appear in the form of $-dt^2 + dz^2$ (in the Gowdy model, multiplied by a conformal factor). While we will not try to explore this kind of idea further here, perhaps theorems about the existence of solutions found by means of the method could be based on this fact.

2.3 Topological Questions

An interesting feature of cosmological models generated by means of horizon and causal structure methods is that for philosophical reasons one often wants these models to have compact $t =$ const. sections, that is, they should be closed cosmological models. This can be achieved for many of the solutions given above by simply specifying the global topology without changing the local geometry. This has caused some problems with these models, especially the vacuum models generated by horizon methods, since they represent incomplete manifolds, and when one passes through the horizon, the spacelike direction in which the manifold would have to be closed becomes timelike, leading to the possibility of closed timelike lines in the "black hole" sector.

Perhaps the simplest model where this occurs is the Kantowski-Sachs-Schwarzschild manifold. If we use (13) for the Kantowski-Sachs model, notice that the $\theta\varphi$ sector has a natural two-sphere topology, and there is no obstruction to compactifying in the r direction by simply assuming that $0 \leq r \leq 2\pi$, with $r = 0$ and $r = 2\pi$ identified, giving the manifold an $S^1 \times S^2$ topology. Of course, if we consider the other side of the horizon, the $t \leftrightarrow r$ reparametrization means that the r-direction becomes the t-direction, and we have $0 \leq t \leq 2\pi$ and the possibility of closed timelike lines. In principle one could have timelike lines which pass from the cosmological region to the black hole region and remain trapped in that region in an eternal closed timelike curve. In the case of the Kantowski- Sachs-Schwarzschild manifold it seems to be impossible for this to happen. From Fig. 4 one can see that timelike geodesics

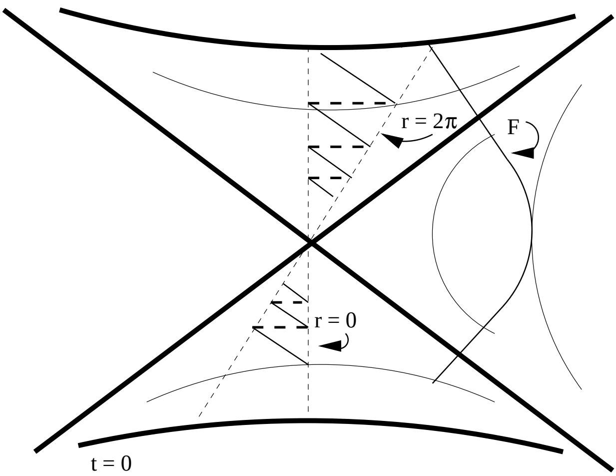


Fig. 4. A Kruskal diagram for geodesics in Kantowski-Sachs-Schwarzschild. For an open topology the timelike geodesic F can leave the cosmological region, enter the Schwarzschild region and return to the cosmological region. This behavior is similar to that of some geodesics in Taub-NUT. For the closed topology, however, geodesics wrap around between $r = 0$ and $r = 2\pi$ (r being the new "radial" direction inside the horizon), following the solid-dashed sawtooth path (here, as an example, for a lightlike geodesic), which always stays inside the horizon except at the crossing point of the past and future horizons

which begin inside the horizon will always stay within the horizon, simply passing through the crossing point of the past and future horizons (a focussing point for all geodesics) [55]. The choice of topology of this metric has a long history. The original studies of this metric assumed that r ran from $-\infty$ to $+\infty$ and that the topology was $R^1 \times S^2$. This topology was usually used until Laflamme and Shellard introduced the $S^1 \times S^2$ topology [56]. Since that time most authors have used it, but some still prefer the $R^1 \times S^2$ topology [57].

Probably the first solution where topological problems were noticed was the Taub-NUT model, where the natural topology of t = const. slices in the

Taub model is S^3, which means that the same topology must apply to the $t\theta\varphi$ sector of NUT space. Here it is possible for timelike geodesics to leave the Taub region and pass into the NUT region, where they may be trapped in trajectories that return to the same spacetime point [45,49,58].

For the other solutions generated by horizon methods mentioned so far, it seems that we will have similar problems, but there has been no study of the effect of topology changes on the causal properties of the solutions. Even degenerate solutions such as the deSitter model and Minkowski space can be given different topologies by means of this type of identifications. Thus, for example, the Milne model, while it is locally flat, may be made into a more acceptable cosmology by means of such a global topology change, but it would suffer from the same problem as the other metrics of this type in that the Rindler space sector would then have an unacceptable topology. The Kerr-Gowdy solution has this same topological difficulty, since the interior is supposed to have an $S^1 \times S^2$ topology, and outside the horizon this implies an S^1 topology in the timelike direction. The causal properties of this solution with this topology seem never to have been studied.

For metrics generated by causal structure methods, there is, in principle no obtruction to topology changes, since there is no "black hole" sector where new topologies can lead to causality problem. The Gowdy T^3 model is a good example of this. While we have not specified the ranges of r, ϕ and z in (29), Gowdy, wishing to have a closed topology for the t = const. surfaces in his model, assumed that $0 \leq r, \phi, z \leq 2\pi$ with points at 0 and 2π identified, which gives the model a three-torus topology. Since there is no "black hole" sector for the new manifold, there is no reason not to choose this topology. The only difference that this change introduces is in the boundary conditions that ψ and γ must satisfy. If r, ϕ, and z run from $-\infty$ to $+\infty$ the solutions of Eq. (30) are built up of eigenfunctions of continuous eigenvalues, while in the Gowdy topology the eigenvalues are discrete.

Note that if we were to try to make the $t \leftrightarrow r$ change for this topology, the Einstein-Rosen wave thus generated would be closed in the time direction, but the Gowdy manifold is complete, and there is no reason to concern ourselves with this possibility.

As we mentioned in the Introduction, the somewhat artificial model we gave there may be closed in the space directions without changing the local metric.

3 Mapping Methods

As we mentioned in the Introduction, the solution generating techniques developed in the last few decades have not often been used in the context of cosmological models. However, it is known that these techniques can be applied when the spacetime possesses at least two commuting Killing vector fields. This is exactly the case of the cosmological models under investigation in this paper.

In the original works on solution generating techniques the central idea was to project the spacetime on the two-dimensional hypersurface defined by the Killing vector fields. This projection allows one to reformulate the spacetime metric and the corresponding field equations in such a way that the symmetries of the differential equations can be investigated in a straightforward manner. An alternative but related procedure is to derive the Ernst representation of the field equations where it is possible to apply (with some modifications) known techniques. In the present work we will use this second alternative.

3.1 Ernst Representation of the T^3 Gowdy Models

In this section we will begin by considering the unpolarized T^3 Gowdy model. We will use a slightly different parametrization of these models from that used in (29) in order to make comparisons with Refs. [59]. The unpolarized model has the metric

$$ds^2 = e^{-\lambda/2}e^{\tau/2}(-e^{-2\tau}d\tau^2 + d\theta^2) + e^{-\tau}[e^P(d\sigma + Qd\delta)^2 + e^{-P}d\delta^2] \ , \quad (32)$$

where the functions λ, P and Q depend on the coordinates τ and θ only, and $0 \leq \sigma, \delta, \theta \leq 2\pi$. These spacetimes are characterized by the existence of two commuting Killing vector fields $\eta_1 = \partial/\partial\sigma$ and $\eta_2 = \partial/\partial\delta$. In the special case $Q = 0$, the fields η_1 and η_2 become hypersurface orthogonal to each other and the metric (32) describes the polarized T^3 Gowdy models.

Einstein's vacuum field equations for the T^3 models consist of a set of two second order differential equations for P and Q

$$P_{,\tau\tau} - e^{-2\tau}P_{,\theta\theta} - e^{2P}(Q_{,\tau}^2 - e^{-2\tau}Q_{,\theta}^2) = 0 \ , \quad (33)$$

$$Q_{,\tau\tau} - e^{-2\tau}Q_{,\theta\theta} + 2(P_{,\tau}Q_{,\tau} - e^{-2\tau}P_{,\theta}Q_{,\theta}) = 0 \ , \quad (34)$$

and two first order differential equations for λ,

$$\lambda_{,\tau} = P_{,\tau}^2 + e^{-2\tau} + e^{2P}(Q_{,\tau}^2 + e^{-2\tau}Q_{,\theta}^2) \ , \quad (35)$$

$$\lambda_{,\theta} = 2(P_{,\theta}P_{,\tau} + e^{2\tau}Q_{,\theta}Q_{,\tau}) \ . \quad (36)$$

The set of equations for λ is the equivalent of Eqs. (31) and can be solved by quadratures once P and Q are known.

As we mentioned, the special case of the polarized T^3 model is obtained from the metric (32) just by taking $Q = 0$. The resulting field equations are easier to handle, and a general solution for the main function P can be found by separation of variables. In fact, let us consider $P(\tau,\theta) = T(\tau)\Theta(\theta)$; then, Eq. (33) separates into

$$\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} = -n^2 \ , \qquad \text{and} \qquad \frac{d^2T}{d\tau^2} + n^2 e^{-2\tau} T = 0 \ , \tag{37}$$

where n is the separation constant which in this case has to be an integer in order for the condition $\Theta(\theta + 2\pi) = \Theta(\theta)$ to be satisfied. With this assumption, the general solution for P can be written as an infinite series of the form

$$P = \sum_{n=0}^{\infty} [A_n \cos(n\theta) + B_n \sin(n\theta)][C_n J_0(ne^{-\tau}) + D_n N_0(ne^{-\tau})] \ , \tag{38}$$

where A_n, B_n, C_n and D_n are arbitrary constants. If we want to avoid singularities at $\tau = +\infty$, the constant D_n has to vanish.

The models described by the general ($Q \neq 0$) metric (32) have been used extensively for numerical investigations in classical as well as in minisuperspace quantum gravity [59]. One of the reasons why these investigations have used numerical methods is because it is usually believed that the set of main field equations (33) and (34) is such a complicated system that analytic solutions would be difficult to find. We will show here that it is possible to generate unpolarized solutions ($Q \neq 0$) from a given polarized solution ($Q = 0$) by using modern solution generating techniques. Of course, the new solutions will be particular solutions, and the numerical investigations in Refs. [59] were carried out with the aim of studying the general behavior of the models near a singularity, information that no particular solution can give. However, families of particular solutions can give us clues about how to set up numerical solutions.

We can apply the solution generating techniques by writing the field equations in such a way that the symmetries involved can be derived and understood easily. To this end, we first introduce a new "time" coordinate $t = e^{-\tau}$ and a new function $R = R(t,\theta)$ by means of the equations

$$R_{,t} = te^{2P} Q_{,\theta} \ , \qquad R_{,\theta} = te^{2P} Q_{,t} \ . \tag{39}$$

Then, the field equations (34) and (33) can be expressed as

$$t^2 \left(P_{,tt} + \frac{1}{t} P_{,t} - P_{,\theta\theta} \right) + e^{-2P}(R_{,t}^2 - R_{,\theta}^2) = 0 \ , \tag{40}$$

$$te^P \left(R_{,tt} + \frac{1}{t} R_{,t} - R_{,\theta\theta} \right) - 2[(te^P)_{,t} R_{,t} - (te^P)_{,\theta} R_{,\theta}] = 0 \ . \tag{41}$$

Furthermore, this last equation for R turns out to be identically satisfied if the integrability condition $R_{,t\theta} = R_{,\theta t}$ is fulfilled.

We can now introduce the complex Ernst potential ϵ and the complex gradient operator D as

$$\epsilon = te^P + iR \ , \qquad \text{and} \qquad D = \left(\frac{\partial}{\partial t} \ , \ i\frac{\partial}{\partial \theta} \right) \ , \tag{42}$$

which allow us to write the main field equations in the *Ernst-like representation*

$$Re(\epsilon)\left(D^2\epsilon + \frac{1}{t} Dt\, D\epsilon \right) - (D\epsilon)^2 = 0 \ . \tag{43}$$

It is easy to see that the field equations (40, 41) can be obtained as the real and imaginary part of the Ernst equation (43), respectively.

The importance of this representation is that it also can be derived from a Lagrangian by means of a variational principle. In turn, that Lagrangian may be interpreted as a metric Lagrangian defined in a two dimensional Riemannian space (the potential space), the coordinates of which are the real and imaginary parts of the Ernst potential. Applying the variational principle in the potential space on the metric Lagrangian, one obtains the corresponding geodesic equations, which turn out to be equivalent to the Ernst equation (43). Hence, to investigate the symmetries of the Ernst equation one can study infinitesimal transformations which leave invariant the geodesic equations in the the potential space. In particular, the transformations associated with the Killing vector fields of the metric in the potential space leave the corresponding geodesic equations invariant. One could think of a solution of the Ernst equation as a geodesic in the potential space, and the Killing vectors of the metric as transformations that starting from a given geodesic lead to a different geodesic, i.e., to a different solution of the Ernst equation. This is the basic idea behind some of the known solution generating techniques.

We will now derive a simple but illustrative symmetry of the Ernst equations which allows to generate new solutions. To this end, we introduce a new complex potential $\xi = \xi(t, \theta)$ by means of the relationship

$$\epsilon = \frac{1-\xi}{1+\xi} \ . \tag{44}$$

Then, the Ernst equation (43) transforms into

$$(1 - \xi\xi^*)\left(D^2\xi + \frac{1}{t} Dt\, D\xi \right) + 2\xi^*(D\xi)^2 = 0 \ , \tag{45}$$

where ξ^* represents the complex conjugate potential. Furthermore, we introduce new coordinates x and y by

$$t^2 = c^2(1-x^2)(1-y^2) \ , \qquad \theta = cxy \ , \tag{46}$$

where c is a real constant. In these coordinates, the main field equation (45) can be written in the following form

$$(1-\xi\xi^*)\{[(1-x^2)\xi_{,x}]_{,x} - [(1-y^2)\xi_{,y}]_{,y}\} + 2\xi^*[(1-x^2)\xi_{,x}^2 - (1-y^2)\xi_{,y}^2] = 0 \ , \tag{47}$$

which explicity shows the invariance with respect to the change of coordinates $x \leftrightarrow y$, i.e., if $\xi(x, y)$ is a solution of (47) then $\xi(y, x)$ is also a solution. The simplest solution of Eq. (47) is $\xi^{-1} = x$, so $\xi^{-1} = y$ is also a solution. A linear combination of these two solutions

$$\xi^{-1} = ax + iby \tag{48}$$

turns out to also be a solution if the condition $a^2 + b^2 = 1$ is satisfied. In this way it is relatively easy to generate new solutions. Once the potential ξ is known one can calculate the functions P and R algebraically by means of Equations (44) and (42). Then, Eq. (39) may be integrated to yield Q. Finally, the solution for the function λ may be calculated by quadratures.

It is interesting to note that one can derive some of the main properties of the Gowdy model just by inspecting the explicit form of the Ernst potential. If the Ernst potential ϵ is real, then the function R vanishes and, therefore, the metric function Q is a constant which can be absorbed by means of a suitable coordinate transformation. Thus, a real Ernst potential corresponds to a polarized Gowdy model. In the case of complex Ernst potential with nontrivial imaginary part (not proportional to its real part), it is guaranteed that the resulting metric corresponds to an unpolarized Gowdy model ($Q \neq 0$). For instance, even the simple solution (48) will lead to a nontrivial unpolarized solution with $Q \neq 0$.

3.2 Generation of New Solutions

In this section we will briefly describe the way new solutions can be generated by using HKX transformations. Let us assume that a solution for the *polarized* Gowdy model is given by $P = P_0(t, \theta)$ and $\lambda = \lambda_0(t, \theta)$. The corresponding Ernst potential is real, and we denote it by $\epsilon_0 = te^{P_0}$.

A HKX transformation acting on ϵ_0 generates a new complex Ernst potential ϵ which will correspond to an unpolarized Gowdy model only if its imaginary part is not proportional to its real part. This will depend on the explicit form of the initial Ernst potential ϵ_0. However, if we apply two different HKX transformations it can be shown that the resulting Ernst potential is nontrivial. For this reason, we present here the result of the action of two HKX transformations on the real Ernst potential ϵ_0. For the sake of simplicity, we use the coordinates x and y defined in Eq. (46) in which the new Ernst potential can be written as [42]

$$\epsilon = \epsilon_0 \frac{x(1 - \mu_1\mu_2) + iy(\mu_1 + \mu_2) + (1 + \mu_1\mu_2) + i(\mu_1 - \mu_2)}{x(1 - \mu_1\mu_2) + iy(\mu_1 + \mu_2) + (1 + \mu_1\mu_2) - i(\mu_1 - \mu_2)}, \tag{49}$$

where we have introduced new functions μ_1 and μ_2 defined by

$$\mu_1 = \alpha_1 e^{2\beta_-} \qquad \text{and} \qquad \mu_2 = \alpha_2 e^{2\beta_+} . \tag{50}$$

Here α_1 and α_2 are real constants and $\beta_\pm$ are functions which satisfy a set of two first order differential equations

$$(x \mp y)(\beta_\pm)_{,x} = (1 \mp xy)\frac{\epsilon_{0,x}}{\epsilon_0} \mp (1 - y^2)\frac{\epsilon_{0,y}}{\epsilon_0} \ , \tag{51}$$

$$(x \mp y)(\beta_\pm)_{,y} = (1 \mp xy)\frac{\epsilon_{0,x}}{\epsilon_0} \pm (1 - x^2)\frac{\epsilon_{0,y}}{\epsilon_0} \ . \tag{52}$$

Thus, the generation of new solutions reduces to the integration of the differential equations for $\beta_\pm$. The new constants α_1 and α_2 have been introduced by the two HKX transformations. For vanishing α_1 and α_2, the potential (49) reduces to the original Ernst potential of the polarized Gowdy model.

Equations (51) and (52) allow us to generate new *unpolarized* solutions starting from a given *polarized* Gowdy model. The derivation of explicit solutions implies the integration of a set of two first order differential equations for a function which determines the Ernst potential of the new solution. In turn, from the Ernst potential one can obtain the new metric functions which completely determine the spacetime of the new cosmological model. In order to complete this procedure it is necessary to carry out lenghtly but straightforward calculations. We are attempting to calculate the unpolarized solution corresponding to the general solution (38) of the polarized T^3 Gowdy model.

4 Conclusions and Suggestions for Further Research

We have given several examples of the three methods listed in the Introduction. It is obvious that these three approaches are far from exhausted, and may give a number of interesting exact solutions to known types of cosmological models and suggest new types of models of interest. It is probable that the horizon and mapping methods may be more productive than causal structure methods.

It seems that horizon methods have barely touched the surface of possible solutions. There are a large number of exact solutions of "black hole" type given in Ref. [46] that are candidates for generating cosmological models inside their horizons. We have looked at the Kantowski-Sachs-Schwarzschild, Taub-NUT and Kerr manifolds in some detail. We have also mentioned as possibilities the Tomimatsu-Sato and Quevedo class of solutions which should lead to models with mixed singularities, part null surface and part curvature singularity. There are large classes of vacuum type D solutions that might give interesting cosmologies [46]. Even such well known solutions as Kerr-Taub-NUT do not seem to have been investigated thoroughly as cosmological models.

Models with matter seem hardly to have been touched. While fluid models may be difficult to generate, the example of electrovac universes shows us that there are many possible cosmologies of this type. Even the interior of the

Reissner-Nordstrøm solution does not seem to have been investigated in this context. It would also be interesting to investigate the possibility of obtaining useful cosmological models from the most general type D electrovac solutions with cosmological constant, found by Debever et al. [61], which contains 13 different parameters and includes the Kerr-Newman solution as a special case. Even the Kerr-Newman solution [13] would generate an $R^1 \times S^2$ (and even, perhaps, an $S^1 \times S^2$) Gowdy solution, equivalent to, or a generalization of, the solutions of Carmeli, Charach and Malin [60] for T^3 models. Other types of solutions, such as scalar field models do not seem to have been investigated at all in this context.

If anything, mapping methods seem even more underutilized in cosmology. The example we have seen is one of many that come to mind. For the Gowdy models one could use these methods to investigate the $S^1 \times S^2$ models and perhaps generate unkown solutions. The corresponding Ernst potential of the field equations present some technical difficulties related to the specific topology of the model. This problem is currently under investigation. For this case, it would also be interesting to generate new unpolarized solutions starting from the general polarized solution, which in the case of $S^1 \times S^2$ models, just as in the T^3 models, can also be represented as an infinite series of eigenfunctions since the corresponding field equations reduce to separable linear second order differential equations.

Solutions with with true curvature singularities on possible inner and outer horizons might be especially interesting. An example of this idea was given by Moncrief [62], who used the Geroch group to generate a solution that had a curvature singularity in place of a horizon in the Kerr-Taub-NUT case.

There will certainly be many opportunities to apply all three of the techniques discussed in the article to the generation of new cosmologies.

Acknowledgments

We wish to thank T. Jacobson, V. Moncrief and O. Obregon for useful discussions. This work was supported in part by CONACyT grant 3567E, and DGAPA-UNAM grants 121298 and IN106097.

References

1. Krasiński A. (1997) Inhomogeneous Cosmological Models. Cambridge U. P., Cambridge.
2. The literature on the deSitter metric is vast. For the description given here see: Hawking S., Ellis G. (1973) The Large Scale Structure of Space-Time. Cambridge U. P., Cambridge; Møller C. (1952) The General Theory of Relativity. Oxford U. P., Oxford.
3. deSitter W. (1917) Mon. Not. Roy. Astron. Soc. **78**, 3.

4. Lemaitre G. (1925) J. Math. and Phys. (M.I.T.) **4**, 188.
5. Robertson H. (1928) Phil. Mag. **5**, 835.
6. Gowdy R. (1971) Phys. Rev. Lett. **27**, 826; (1974) Ann. Phys. (N.Y.) **83**, 203.
7. Bondi H., Pirani F., Robinson I. (1959) Proc. Roy. Soc. London **A251**, 519; Ehlers J., Kundt W., (1962) in: Witten L. (Ed.) Gravitation: An Introduction to Current Research. Wiley, New York.
8. Kerr R. P. (1963) Phys. Rev. Lett. **11**, 237.
9. Newman E. T., Tamburino L. and Unti T. (1963) J. Math. Phys. **4**, 915.
10. Kinnersley W. (1969) J. Math. Phys. **10**, 1195.
11. Plebański J. F. and Demiański M. (1976) Ann. Phys.(N. Y.) **98**, 98.
12. Newman E. T. and Janis A. I. (1965) J. Math. Phys. **6**, 915.
13. Newman E. T. *et al.* (1965) J. Math. Phys. **6**, 918.
14. Demiański M. and Newman E. T. (1966) Bull. Acad. Pol. Sci., Ser. Math. Astron. Phys. **14**, 653.
15. Carter B. (1968) Comm. Math. Phys. **10**, 280.
16. Demiański M. (1973) Acta Astronom. Pol. **23**, 197.
17. Ernst F. J. (1968) Phys. Rev. **167**, 1175.
18. Ernst F. J. (1968) Phys. Rev. **168**, 1415.
19. Kinnersley W. (1973) J. Math.. Phys. **14**, 651.
20. Kinnersley W. (1977) J. Math. Phys. **18**, 1529.
21. Tomimatsu A. and Sato H. (1972) Phys. Rev. Lett. **29**, 1344.
22. Tomimatsu A. and Sato H. (1973) Prog. Theor. Phys. **50**, 95.
23. Yamazaki M. and Hori S. (1977) Prog. Theor. Phys. (Lett.) **57**, L696.
24. Yamazaki M. (1977) Prog. Theor. Phys. **57**, 1951.
25. Yamazaki M. (1977) J. Math. Phys. **18**, 2502.
26. Geroch R. (1971) J. Math. Phys. **12**, 918.
27. Geroch R. (1972) J. Math. Phys. **13**, 394.
28. Cosgrove C. M. (1977) J. Phys. A: Math. Gen. **10**, 1481.
29. Cosgrove C. M. (1977) J. Phys. A: Math. Gen. **10**, 2093.
30. Kinnersley W. and Chitre D. M. (1977) J. Math. Phys. **18**, 1538.
31. Kinnersley W. and Chitre D. M. (1978) J. Math. Phys. **19**, 1926.
32. Kinnersley W. and Chitre D. M. (1978) J. Math. Phys. **19**, 2037.
33. Kinnersley W. and Chitre D. M. (1977) Phys. Rev. Lett. **40**, 1608.
34. Hoenselaers C., Kinnersley W., and Xanthopoulos B. C. (1979) Phys. Rev. Lett. **42**, 481.
35. Hoenselaers C., Kinnersley W., and Xanthopoulos B. C. (1979) J. Math. Phys. **20**, 2530.
36. Harrison B. K. (1978) Phys. Rev. Lett. **41**, 1197.
37. Neugebauer G. (1980) J. Phys. A: Math. Gen. **13**, 1737, L19.
38. Belinsky V. A. and Zakharov V. E. (1978) Soviet Phys.-JETP **48**, 985.
39. Belinsky V. A. and Zakharov V. E. (1978) Soviet Phys.-JETP **50**, 1.
40. Cosgrove C. M. (1980) J. Math. Phys. **21**, 2417.
41. Cosgrove C. M. (1982) in: Proceedings of the Second Marcel Grossmann Meeting on General Relativity, Ruffini R. (Ed.), North-Holland, Amsterdam.
42. Quevedo H. (1990) Fortschr. Phys. **38**, 10.
43. See, for example, Rindler W. (1969) Essential Relativity: Special, General, and Cosmological (Second Ed.). Springer-Verlag, Heidelberg.
44. Taub A. (1951) Ann. Math. **53**, 472.

45. Misner C. (1963) in: Relativity Theory and Astrophysics 1. Ehlers J. (Ed.), American Mathematical Society, Providence R. I.; Misner C., Taub A. (1968) Zh. Eksp. Teor. Fiz. **55**, 233 (Soviet Physics – JETP **28**, 122 [1969]).
46. Kramer D., Stephani D., Herlt E., MacCallum M., Schmutzer E. (1980) Exact Solutions of Einstein's Field Equations. Cambridge U. P., Cambridge.
47. See Burko L., Ori A. (Eds.) (1997) Proceedings of the Workshop on the Internal Structure of Black Holes and Spacetime Singularities. IOP, Bristol.
48. Obregón O., Ryan M. (1998) gr-qc/9810068
49. Ryan M., Shepley L. (1975) Homogeneous Relativistic Cosmologies. Princeton U. P., Princeton.
50. Conradi H-D. (1995) Class. Q. Grav. **12**, 2423.
51. Brill D. (1964) Phys. Rev. **133**, B845.
52. Israel W. (1967) Phys. Rev. **164**, 1776; Carter B. (1970) Phys. Rev. Lett. **26**, 331; Hawking S. (1972) Commun. Math. Phys. **25**, 152.
53. Quevedo H. (1986) Phys. Rev. D **33**, 324.
54. Einstein A., Rosen N. (1937) J. Franklin Inst. **223**, 43.
55. Jacobson T. (1998) Private communication.
56. Laflamme R., Shellard E. P. S. (1987) Phys. Rev. D **35**, 2315.
57. Xanthopolous B., Zannias T. (1992) J. Math. Phys. **33**, 1415; (1991) J. Math. Phys. **30**, 1121.
58. Hawking S., Ellis G. in Ref. [2].
59. See Berger B., Moncrief V. (1993) Phys. Rev. D **48**, 4676; Berger B., Garfinkle D. (1998) Phys. Rev. D **57**, 4767, and references therein.
60. Charach Ch. (1979) Phys. Rev. D **19**, 1058; Phys. Rev. D **19**, 3516; Carmeli M., Charach Ch., Malin S. (1981) Phys. Reports **76**, 1.
61. Debever R., Kamran N., McLenaghan R. G. (1984) J. Math. Phys. **25**, 1955.
62. Moncrief V. (1987) Class. Q. Grav. **6**, 1555.

Multidimensional Cosmological and Spherically Symmetric Solutions with Intersecting *p*-Branes

V.D. Ivashchuk and V.N. Melnikov

Center for Gravitation and Fundamental Metrology VNIIMS, 3-1 M. Ulyanovoy Str., Moscow, 117313, Russia
e-mail: melnikov@rgs.phys.msu.su

Abstract. Multidimensional model describing the cosmological evolution and/ or spherically symmetric configuration with $(n+1)$ Einstein spaces in the theory with several scalar fields and forms is considered. When electro-magnetic composite p-brane ansatz is adopted, n "internal" spaces are Ricci-flat, one space M_0 has a non-zero curvature, and all p-branes do not "live" in M_0, a class of exact solutions is obtained if certain block-orthogonality relations on p-brane vectors are imposed. A subclass of spherically-symmetric solutions containing non-extremal p-brane black holes is considered. Post-Newtonian parameters are calculated and some examples are considered.
Keywords. P-branes, multidimensional cosmology, black holes.

1 Introduction

The necessity of studying multidimensional models of gravitation and cosmology [1,2] is motivated by several reasons. First, the main trend of modern physics is the unification of all known fundamental physical interactions: electromagnetic, weak, strong and gravitational ones. During last decades there was a significant progress in unifying weak and electromagnetic interactions, some more modest achievements in GUT, supersymmetric, string and superstring theories.

Now theories with membranes, p-branes and more vague M- and F-theories [4,6,7,8] are being created and studied. Having no any definite successful theory of unification now, it is desirable to study the common features of these theories and their applications to solving basic problems of modern gravity and cosmology. Moreover, if we really believe in unified theories, the early stages of the Universe evolution, as a unique superhigh energy region, is the most proper and natural arena for them.

Second, multidimensional gravitational models, as well as scalar-tensor theories of gravity, are the theoretical framework for describing possible temporal and range variations of fundamental physical constants [3]. These ideas originated from earlier papers of P.Dirac (1937) on relations between phenomena of micro and macro worlds and up till now they are under a thorough study both theoretically and experimentally.

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 214–247, 2000.

At last, applying multidimensional gravitational models to basic problems of modern cosmology and black hole physics we hope to find answers to such long standing problems as singular or nonsingular initial states, creation of the Universe, creation of matter and entropy in it, origin of inflation and specific scalar fields which are necessary for its realization, isotropization and graceful exit problems, stability and nature of fundamental constants [10], possible number of extra dimensions, their stable compactification etc.

Bearing in mind that multidimensional gravitational models are certain generalizations of general relativity which is tested reliably for weak fields up to 0,001 (they may be viewed as some effective scalar-tensor theories in simple variants in four dimensions) it is quite natural to inquire about their possible observational or experimental windows. From what we already know, among these windows are:

- possible deviations from the Newton and Coulomb laws,
- possible variations of the effective gravitational constant with a time rate less than the Hubble one,
- possible existence of monopole modes in gravitational waves,
- different behaviour of strong field objects, such as multidimensional black holes, wormholes and p-branes,
- standard cosmological tests etc.

As modern cosmology already became a unique laboratory for testing standard unified models of physical interactions at energies that are far beyond the level of existing and future man-made accelerators and other installations on Earth, there exists a possibility of using cosmological and astrophysical data for discriminating between future unified schemes.

As no accepted unified model exists, in our approach we adopt simple, but general from the point of view of number of dimensions, models based on multidimensional Einstein equations with or without sources of different nature:

- cosmological constant,
- perfect and viscous fluids,
- scalar and electromagnetic fields,
- plus their interactions,
- fields of antisymmetric forms (related to p-branes) etc.

Our main objective was and is to obtain exact solutions (integrable models) for these model self-consistent systems and then to analyze them in cosmological, spherically and axially symmetric cases. In our view this is a natural and most reliable way to study highly nonlinear systems. It is done mainly within the Riemannian geometry. Some simple models in integrable Weyle geometry and with torsion were studied also.

1.1 Problem of Stability of G

Absolute G Measurements The value of the Newton's gravitational constant G as adopted by CODATA in 1986 is based on Luther and Towler measurements of 1982.

Even at that time other existing on 100ppm level measurements deviated from this value more than their uncertainties [12]. During last years the situation, after very precise measurements of G in Germany and New Zealand, became much more vague. Their results deviate from the official CODATA value from 600 ppm at minimal to 630 ppm at maximal values.

As it is seen from the most recent data announced in November 1998 at the Cavendish conference in London the situation with terrestrail absolute G measurements is not improving. The reported values for G (in units of 10^{11}) and their estimated error in ppm are as follows:

Fitzgerald and Armstrong	6.6742	90 ppm
	6.6746	134
Nolting et al. (Zurich)	6.6749	210
Meyer et al. (Wuppethal)	6.6735	240
Karagioz et al. (Moscow)	6.6729	75
Richman et al.	6.683	1700
Schwarz et al.	6.6873	1400
CODATA (1986, Luther)	6.67259	128

This means that either the limit of terrestrial accuracies is reached or we have some new physics entering the measurement procedure [13,14]. First means that we should shift to space experiments to measure G [15] and second means that more thorough study of theories generalizing Einstein's general relativity is necessary.

Data on Temporal Variations of G Dirac's prediction based on his Large Numbers Hypothesis is $\dot{G}/G = (-5)10^{-11}\ year^{-1}$. Other hypotheses and theories, in particular some scalar-tensor or multidimensional ones, predict these variations on the level of $10^{-12} - 10^{-13}$ per year. As to experimental or observational data, the results are rather nonconclusive. The most reliable ones are based on Mars orbiters and landers (Hellings,1983) and on lunar laser ranging (Muller et al., 1993; Williams et al., 1996). They are not better than 10^{-12} per year [16]. Here once more we see that there is a need for corresponding theoretical and experimental studies. Probably, future space missions to other planets will be a decisive step in solving the problem of temporal variations of G and defining the fates of different theories which predict them as the larger is the time interval between successive measurements and, of course, the more precise they are, the more stringent results will be obtained.

Non-Newtonian Interactions (EP and ISL Tests) Nearly all modified theories of gravity and unified theories predict also some deviations from the

Newton law (ISL) or composite-dependant violation of the Equivalence Principle (EP) due to an appearance of new possible massive particles (partners) [11]. Experimental data exclude the existence of these particles nearly at all ranges except less than millimeter and also at meters and hundreds of meters ranges. The most recent result in the range of 20-500 m was obtained by Achilli et al [17]. They found the positive result for the deviation from the Newton law with the Yukawa potential strength alpha between 0,13 and 0,25. Of course, these results need to be verified in other independent experiments, probably in space ones.

1.2 Multidimensional Models

The history of multidimensional approach starts from the well-known papers of T.K. Kaluza and O. Klein [18,19] on 5-dimensional theories which opened an interest (see [20,21,22,23]) to investigations in multidimensional gravity. These ideas were continued by P.Jordan [24] who suggested to consider the more general case $g_{55} \neq$ const leading to the theory with an additional scalar field. The papers [18,19,24] were in some sense a source of inspiration for C. Brans and R.H. Dicke in their well-known work on the scalar-tensor gravitational theory [25]. After their work a lot of investigations were done using material or fundamental scalar fields, both conformal and nonconformal (see details in [3]).

The revival of ideas of many dimensions started in 70th and continues now. It is due completely to the development of unified theories. In the 70th an interest to multidimensional gravitational models was stimulated mainly by: i) the ideas of gauge theories leading to the non-Abelian generalization of Kaluza-Klein approach and by ii) supergravitational theories [26,27]. In the 80th the supergravitational theories were "replaced" by superstring models [28]. Now it is heated by expectations connected with overall M-theory or even some F-theory. In all these theories 4-dimensional gravitational models with extra fields were obtained from some multidimensional model by a dimensional reduction based on the decomposition of the manifold

$$M = M^4 \times M_{int},$$

where M^4 is a 4-dimensional manifold and M_{int} is some internal manifold (mostly considered as a compact one).

The earlier papers on multidimensional cosmology dealt with multidimensional Einstein equations and with a block-diagonal cosmological metric defined on the manifold $M = \mathbb{R} \times M_0 \times ... \times M_n$ of the form

$$g = -dt \otimes dt + \sum_{r=0}^{n} a_r^2(t) g^r$$

where (M_r, g^r) are Einstein spaces, $r = 0, \ldots, n$ [30]–[62]. In some of them a cosmological constant and simple scalar fields were used also [104].

In [40,41,45,50,51,57,69,70] the models with higher dimensional "perfect-fluid" were considered. In these models pressures (for any component) are proportional to the density

$$p_r = \left(1 - \frac{u_r}{d_r}\right)\rho,$$

$r = 0, \dots, n$, where d_r is a dimension of M_r. Such models are reduced to pseudo-Euclidean Toda-like systems with the Lagrangian

$$L = \frac{1}{2} G_{ij} \dot{x}^i \dot{x}^j - \sum_{k=1}^{m} A_k \mathrm{e}^{u_i^k x^i}$$

and the zero-energy constraint $E = 0$. In a classical case exact solutions with Ricci-flat (M_r, g^r) for 1-component case were considered by many authors (see, for example, [38,39,50,51,69,70,97] and references therein). For the two component perfect-fluid there were solutions with two curvatures, i.e. $n = 2$, when $(d_1, d_2) = (2,8), (3,6), (5,5)$ [106] and corresponding non-singular solutions from [141]. Among the solutions [106] there exists a special class of Milne-type solutions. Recently some interesting extensions of 2-component solutions were obtained in [107].

It should be noted that the pseudo-Euclidean Toda-like systems are not well-studied yet. There exists a special class of equations of state that gives rise to the Euclidean Toda models. First such solution was considered in [70] for the Lie algebra a_2. Recently the case of $a_n = sl(n+1)$ Lie algebras was considered and the solutions were expressed in terms of a new elegant representation (obtained by Anderson) [105].

The cosmological solutions may have regimes with: i) spontaneous and dynamical compactifications; ii) Kasner-like and billiard behavior near the singularity; iii) inflation and izotropization for large times (see, for example, [104,69]).

Near the singularity one can have an oscillating behavior like in the well-known mixmaster (Bianchi-IX) model. Multidimensional generalizations of this model were considered by many authors (see, for example, [29,76,77,78]). In [79,80,81] the billiard representation for multidimensional cosmological models near the singularity was considered and the criterion for the volume of the billiard to be finite was established in terms of illumination of the unit sphere by point-like sources. For perfect-fluid this was considered in detail in [81]. Some interesting topics related to general (non-homogeneous) situation were considered in [82].

Multidimensional cosmological models have a generalization to the case when the bulk and shear viscosity of the "fluid" is taken into account [108]. Some classes of exact solutions were obtained, in particular nonsingular cosmological solutions, generation of mass and entropy in the Universe.

Multidimensional quantum cosmology based on the Wheeler-DeWitt (WDW) equation

$$\hat{H}\Psi = 0,$$

where Ψ is the so-called "wave function of the universe", was treated first in [52], (see also [64]). This equation was considered for the vacuum case in [52] and integrated in a very special situation of 2-spaces. The WDW equation for the cosmological constant and for the "perfect-fluid" was investigated in [68,104] and [69] respectively.

Exact solutions in 1-component case were considered in detail in [97] (for perfect fluid). In [60] the multidimensional quantum wormholes were suggested, i.e. solutions with a special-type behavior of the wave function (see [65]).

These solutions were generalized to account for cosmological constant in [68,104] and to the perfect-fluid case in [69,97]. In [81] the "quantum billiard" was obtained for multidimensional WDW solutions near the singularity. It should be also noted that the "third-quantized" multidimensional cosmological models were considered in several papers [59,103,97]. One may point out that in all cases when we had classical cosmological solutions in many dimensions, the corresponding quantum cosmological solutions were found also.

Cosmological solutions are closely related to solutions with the spherical symmetry. Moreover, the scheme of obtaining them is very similar to the cosmological approach. The first multidimensional generalization of such type was considered by D. Kramer [87] and rediscovered by A.I. Legkii [88], D.J. Gross and M.J. Perry [89] (and also by Davidson and Owen). In [91] the Schwarzschild solution was generalized to the case of n internal Ricci-flat spaces and it was shown that black hole configuration takes place when scale factors of internal spaces are constants. In [92] an analogous generalization of the Tangherlini solution [90] was obtained. These solutions were also generalized to the electrovacuum case [93,96,94]. In [95,94] multidimensional dilatonic black holes were singled out. An interesting theorem was proved in [94] that "cuts" all non-black-hole configurations as non-stable under even monopole perturbations. In [98] the extremely-charged dilatonic black hole solution was generalized to multicenter (Majumdar-Papapetrou) case when the cosmological constant is non-zero.

We note that for $D = 4$ the pioneering Majumdar-Papapetrou solutions with conformal scalar field and electromagnetic field were considered in [161].

At present there exists a special interest to the so-called M- and F-theories etc. [4,6,7,8]. These theories are "supermembrane" analogues of superstring models [28] in $D = 11, 12$ etc. The low-energy limit of these theories leads to models governed by the Lagrangian

$$\mathcal{L} = R[g] - h_{\alpha\beta} g^{MN} \partial_M \varphi^\alpha \partial_N \varphi^\beta - \sum_{a \in \Delta} \frac{\theta_a}{n_a!} \exp[2\lambda_a(\varphi)](F^a)^2,$$

where g is metric, $F^a = dA^a$ are forms of rank $F^a = n_a$, and φ^α are scalar fields.

In [143] it was shown that after dimensional reduction on the manifold $M_0 \times M_1 \times \ldots \times M_n$ and when the composite p-brane ansatz is considered the problem is reduced to the gravitating self-interacting σ-model with certain constraints imposed. For electric p-branes see also [141,142,144] (in [144] the composite electric case was considered). This representation may be considered as a powerful tool for obtaining different solutions with intersecting p-branes (analogs of membranes). In [143,168] the Majumdar-Papapetrou type solutions were obtained (for non-composite electric case see [141,142] and for composite electric case see [144]). These solutions correspond to Ricci-flat (M_i, g^i), $i = 1, \ldots, n$, and were generalized also to the case of Einstein internal spaces [143]. Earlier some special classes of these solutions were considered in [126,127,128,146,147,148]. The obtained solutions take place, when certain orthogonality relations (on couplings parameters, dimensions of "branes", total dimension) are imposed. In this situation a class of cosmological and spherically-symmetric solutions was obtained [166]. Special cases were also considered in [130,153,154,152]. The solutions with the horizon were considered in details in [131,149,150,151,166]. In [151,167] some propositions related to i) interconnection between the Hawking temperature and the singularity behaviour, and ii) to multitemporal configurations were proved.

It should be noted that multidimensional and multitemporal generalizations of the Schwarzschild and Tangherlini solutions were considered in [96, 164], where the generalized Newton's formulas in multitemporal case were obtained.

We note also that there exists a large variety of Toda solutions (open or closed) when certain intersection rules are satisfied [166].

In [166] (see also [158]) the Wheeler-DeWitt equation was integrated for intersecting p-branes in orthogonal case and corresponding classical solutions were obtained also. A slightly different approach was suggested in [155]. (For non-composite case see also [154].)

In [159,169] exact solutions for multidimensional models with intersecting p-branes in case of static internal spaces were obtained. They turned to be de Sitter or anti- de Sitter type. Generation of the effective cosmological constant and inflation via p-branes was demonstrated there. These solutions may be considered as an interesting first step for a quantum description of low-energy limits in different super-p-branes theories.

In this paper we continue our investigations of p-brane solutions (see for example [112,113,148] and references therein) based on sigma-model approach [143,142,144] . (For pure gravitational sector see [109,141].)

Here we consider a cosmological and/or spherically symmetric case , when all functions depend upon one variable (time or radial variable). The model under consideration contains several scalar fields and antisymmetric forms and is governed by action (2.1).

The considered cosmological model contains some stringy cosmological models (see for example [156]). It may be obtained (at classical level) from multidimensional cosmological model with perfect fluid [69,70] as a special case.

Here we find a family of solutions depending on one variable describing the (cosmological or spherically symmetric) "evolution" of $(n+1)$ Einstein spaces in the theory with several scalar fields and forms. When an electromagnetic composite p-brane ansatz is adopted the field equations are reduced to the equations for Toda-like system.

In the case when n "internal" spaces are Ricci-flat, one space M_0 has a non-zero curvature, and all p-branes do not "live" in M_0, we find a family of solutions (Section 4) to the equations of motion (equivalent to equations for Toda-like Lagrangian with zero-energy constraint [166]) if certain block-orthogonality relations on p-brane vectors U^s are imposed. These solutions generalize the solutions from [166] with orthogonal set of vectors U^s. A special class of "block-orthogonal" solutions (with coinciding parameters ν_s inside blocks) was considered earlier in [167].

Here we consider a subclass of spherically-symmetric solutions (Sect. 5). This subclass contains non-extremal p-brane black holes for zero values of "Kasner-like" parameters. The relation for the Hawking temperature is presented (in the black hole case).

We also calculate Post-Newtonian parameters β and γ (Eddington parameters) for the spherically-symmetric solutions (Sect. 6). These parameters may be useful for possible physical applications.

2 The Model

Here like in [143] we consider the model governed by the action

$$\begin{aligned} S = \ & \frac{1}{2\kappa^2} \int_M d^D z \sqrt{|g|} \{ R[g] - 2\Lambda - h_{\alpha\beta}\, g^{MN} \partial_M \varphi^\alpha \partial_N \varphi^\beta \\ & - \sum_{a \in \Delta} \frac{\theta_a}{n_a!} \exp[2\lambda_a(\varphi)] (F^a)^2_g \} + S_{GH}, \end{aligned} \tag{2.1}$$

where $g = g_{MN} dz^M \otimes dz^N$ is the metric $(M, N = 1, \ldots, D)$, $\varphi = (\varphi^\alpha) \in \mathbb{R}^l$ is a vector from dilatonic scalar fields, $(h_{\alpha\beta})$ is a non-degenerate symmetric $l \times l$ matrix $(l \in \mathbb{N})$, $\theta_a = \pm 1$,

$$F^a = dA^a = \frac{1}{n_a!} F^a_{M_1 \ldots M_{n_a}} dz^{M_1} \wedge \ldots \wedge dz^{M_{n_a}} \tag{2.2}$$

is a n_a-form $(n_a \geq 1)$ on a D-dimensional manifold M, Λ is cosmological constant and λ_a is a 1-form on $\mathbb{R}^l$: $\lambda_a(\varphi) = \lambda_{a\alpha} \varphi^\alpha$, $a \in \Delta$, $\alpha = 1, \ldots, l$. In (2.1) we denote $|g| = |\det(g_{MN})|$,

$$(F^a)^2_g = F^a_{M_1 \ldots M_{n_a}} F^a_{N_1 \ldots N_{n_a}} g^{M_1 N_1} \ldots g^{M_{n_a} N_{n_a}}, \tag{2.3}$$

$a \in \Delta$, where Δ is some finite set, and $S_{\rm GH}$ is the standard Gibbons-Hawking boundary term [163]. In the models with one time all $\theta_a = 1$ when the signature of the metric is $(-1, +1, \ldots, +1)$.

The equations of motion corresponding to (2.1) have the following form

$$R_{MN} - \frac{1}{2} g_{MN} R = T_{MN} - \Lambda g_{MN}, \tag{2.4}$$

$$\triangle[g]\varphi^\alpha - \sum_{a \in \Delta} \theta_a \frac{\lambda^\alpha_a}{n_a!} e^{2\lambda_a(\varphi)} (F^a)^2_g = 0, \tag{2.5}$$

$$\nabla_{M_1}[g](e^{2\lambda_a(\varphi)} F^{a, M_1 \ldots M_{n_a}}) = 0, \tag{2.6}$$

$a \in \Delta$; $\alpha = 1, \ldots, l$. In (2.5) $\lambda^\alpha_a = h^{\alpha\beta}\lambda_{a\beta}$, where $(h^{\alpha\beta})$ is matrix inverse to $(h_{\alpha\beta})$. In (2.4)

$$T_{MN} = T_{MN}[\varphi, g] + \sum_{a \in \Delta} \theta_a e^{2\lambda_a(\varphi)} T_{MN}[F^a, g], \tag{2.7}$$

where

$$T_{MN}[\varphi, g] = h_{\alpha\beta} \left(\partial_M \varphi^\alpha \partial_N \varphi^\beta - \frac{1}{2} g_{MN} \partial_P \varphi^\alpha \partial^P \varphi^\beta \right), \tag{2.8}$$

$$T_{MN}[F^a, g] = \frac{1}{n_a!} \left[-\frac{1}{2} g_{MN} (F^a)^2_g + n_a F^a_{M M_2 \ldots M_{n_a}} F_N^{a, M_2 \ldots M_{n_a}} \right]. \tag{2.9}$$

In (2.5), (2.6) $\triangle[g]$ and $\bigtriangledown[g]$ are Laplace-Beltrami and covariant derivative operators respectively corresponding to g.

Let us consider the manifold

$$M = \mathbb{R} \times M_0 \times \ldots \times M_n \tag{2.10}$$

with the metric

$$g = w \mathrm{e}^{2\gamma(u)} du \otimes du + \sum_{i=0}^{n} \mathrm{e}^{2\phi^i(u)} g^i, \tag{2.11}$$

where $w = \pm 1$, u is a distinguished coordinate which, by convention, will be called "time"; $g^i = g^i_{m_i n_i}(y_i) dy_i^{m_i} \otimes dy_i^{n_i}$ is a metric on M_i satisfying the equation

$$R_{m_i n_i}[g^i] = \xi_i g^i_{m_i n_i}, \tag{2.12}$$

$m_i, n_i = 1, \ldots, d_i$; $d_i = \dim M_i$, $\xi_i = {\rm const}$, $i = 0, \ldots, n$; $n \in \mathbb{N}$. Thus, (M_i, g^i) are Einstein spaces. The functions $\gamma, \phi^i \colon (u_-, u_+) \to \mathbb{R}$ are smooth.

Each manifold M_i is assumed to be oriented and connected, $i = 0, \ldots, n$. Then the volume d_i-form

$$\tau_i = \sqrt{|g^i(y_i)|}\, dy_i^1 \wedge \ldots \wedge dy_i^{d_i}, \tag{2.13}$$

and the signature parameter

$$\varepsilon(i) = {\rm sign} \det(g^i_{m_i n_i}) = \pm 1 \tag{2.14}$$

are correctly defined for all $i = 0, \ldots, n$.

Let

$$\Omega_0 = \{\emptyset, \{0\}, \{1\}, \ldots, \{n\}, \{0,1\}, \ldots, \{0,1,\ldots,n\}\} \tag{2.15}$$

be a set of all subsets of

$$I_0 \equiv \{0, \ldots, n\}. \tag{2.16}$$

Let $I = \{i_1, \ldots, i_k\} \in \Omega_0$, $i_1 < \ldots < i_k$. We define a form

$$\tau(I) \equiv \tau_{i_1} \wedge \ldots \wedge \tau_{i_k}, \tag{2.17}$$

of rank

$$d(I) \equiv \sum_{i \in I} d_i, \tag{2.18}$$

and a corresponding p-brane submanifold

$$M_I \equiv M_{i_1} \times \ldots \times M_{i_k}, \tag{2.19}$$

where $p = d(I) - 1$ ($\mathrm{dim M_I} = d(I)$). We also define ε-symbol

$$\varepsilon(I) \equiv \varepsilon(i_1) \ldots \varepsilon(i_k). \tag{2.20}$$

For $I = \emptyset$ we put $\tau(\emptyset) = \varepsilon(\emptyset) = 1$, $d(\emptyset) = 0$.

For fields of forms we adopt the following "composite electro-magnetic" ansatz

$$F^a = \sum_{I \in \Omega_{a,e}} \mathcal{F}^{(a,e,I)} + \sum_{J \in \Omega_{a,m}} \mathcal{F}^{(a,m,J)}, \tag{2.21}$$

where

$$\mathcal{F}^{(a,e,I)} = d\Phi^{(a,e,I)} \wedge \tau(I), \tag{2.22}$$

$$\mathcal{F}^{(a,m,J)} = \mathrm{e}^{-2\lambda_a(\varphi)} * \left(d\Phi^{(a,m,J)} \wedge \tau(J) \right), \tag{2.23}$$

$a \in \Delta$, $I \in \Omega_{a,e}$, $J \in \Omega_{a,m}$ and

$$\Omega_{a,e}, \Omega_{a,m} \subset \Omega_0. \tag{2.24}$$

(For empty $\Omega_{a,v} = \emptyset$, $v = e, m$, we put $\sum\limits_{\emptyset} = 0$ in (2.21)). In (2.23) $* = *[g]$ is the Hodge operator on (M, g).

For the potentials in (2.22), (2.23) we put

$$\Phi^s = \Phi^s(u), \tag{2.25}$$

$s \in S$, where

$$S = S_e \sqcup S_m, \qquad S_v \equiv \sqcup_{a \in \Delta} \{a\} \times \{v\} \times \Omega_{a,v}, \tag{2.26}$$

$v = e, m$. Here $\sqcup$ means the union of non-intersecting sets. The set S consists of elements $s = (a_s, v_s, I_s)$, where $a_s \in \Delta$, $v_s = e, m$ and $I_s \in \Omega_{a,v_s}$ are "color", "electro-magnetic" and "brane" indices, respectively.

For dilatonic scalar fields we put

$$\varphi^\alpha = \varphi^\alpha(u), \tag{2.27}$$

$\alpha = 1, \dots, l$.

From (2.22) and (2.23) we obtain the relations between dimensions of p-brane worldsheets and ranks of forms

$$d(I) = n_a - 1, \quad I \in \Omega_{a,e}, \tag{2.28}$$

$$d(J) = D - n_a - 1, \quad J \in \Omega_{a,m}, \tag{2.29}$$

in electric and magnetic cases respectively.

3 σ-Model Representation

Here, like in [166], we impose a restriction on p-brane configurations, or, equivalently, on $\Omega_{a,v}$. We assume that the energy momentum tensor (T_{MN}) has a block-diagonal structure (as it takes place for (g_{MN})). Sufficient restrictions on $\Omega_{a,v}$ that guarantee a block-diagonality of (T_{MN}) are presented in Appendix 1.

It follows from [143] (see Proposition 2 in [143]) that the equations of motion (2.4)–(2.6) and the Bianchi identities

$$d\mathcal{F}^s = 0, \quad s \in S \tag{3.1}$$

for the field configuration (2.11), (2.21)–(2.23), 2.25, (2.27) with the restrictions (8.2), (8.3) (from Appendix 1) imposed are equivalent to equations of motion for σ-model with the action

$$\begin{aligned} S_\sigma = \frac{\mu_*}{2} \int du \mathcal{N} \Big\{ & G_{ij} \dot{\phi}^i \dot{\phi}^j + h_{\alpha\beta} \dot{\varphi}^\alpha \dot{\varphi}^\beta \\ & + \sum_{s \in S} \varepsilon_s \exp[-2U^s(\phi, \varphi)] (\dot{\Phi}^s)^2 - 2\mathcal{N}^{-2} V(\phi) \Big\}, \end{aligned} \tag{3.2}$$

where $\dot{x} \equiv dx/du$,

$$V = V(\phi) = -w\Lambda \mathrm{e}^{2\gamma_0(\phi)} + \frac{w}{2} \sum_{i=0}^{n} \xi_i d_i \mathrm{e}^{-2\phi^i + 2\gamma_0(\phi)} \tag{3.3}$$

is the potential with

$$\gamma_0(\phi) \equiv \sum_{i=0}^{n} d_i \phi^i, \tag{3.4}$$

and

$$\mathcal{N} = \exp(\gamma_0 - \gamma) > 0 \tag{3.5}$$

is the lapse function,

$$U^s = U^s(\phi, \varphi) = -\chi_s \lambda_{a_s}(\varphi) + \sum_{i \in I_s} d_i \phi^i, \tag{3.6}$$

$$\varepsilon_s = (-\varepsilon[g])^{(1-\chi_s)/2} \varepsilon(I_s) \theta_{a_s} \tag{3.7}$$

for $s = (a_s, v_s, I_s) \in S$, $\varepsilon[g] = \operatorname{sign}\det(g_{MN})$, (more explicitly (3.7) reads $\varepsilon_s = \varepsilon(I_s)\theta_{a_s}$ for $v_s = e$ and $\varepsilon_s = -\varepsilon[g]\varepsilon(I_s)\theta_{a_s}$, for $v_s = m$)

$$\chi_s = +1, \quad v_s = e; \tag{3.8}$$

$$\chi_s = -1, \quad v_s = m, \tag{3.9}$$

and

$$G_{ij} = d_i \delta_{ij} - d_i d_j \tag{3.10}$$

are components of the "pure cosmological" minisupermetric; $i, j = 0, \ldots, n$ [52].

In the electric case ($\mathcal{F}^{(a,m,I)} = 0$) for finite internal space volumes V_i the action (3.2) coincides with the action (2.1) if $\mu_* = -w/\kappa_0^2$, $\kappa^2 = \kappa_0^2 V_0 \ldots V_n$.

Action (3.2) may be also written in the form

$$S_\sigma = \frac{\mu_*}{2} \int du \mathcal{N} \left\{ \mathcal{G}_{\hat{A}\hat{B}}(X) \dot{X}^{\hat{A}} \dot{X}^{\hat{B}} - 2\mathcal{N}^{-2} V(X) \right\}, \tag{3.11}$$

where $X = (X^{\hat{A}}) = (\phi^i, \varphi^\alpha, \Phi^s) \in \mathbf{R}^N$, and minisupermetric

$$\mathcal{G} = \mathcal{G}_{\hat{A}\hat{B}}(X) dX^{\hat{A}} \otimes dX^{\hat{B}} \tag{3.12}$$

on minisuperspace

$$\mathcal{M} = \mathbb{R}^N, \quad N = n + 1 + l + |S| \tag{3.13}$$

($|S|$ is the number of elements in S) is defined by the relation

$$(\mathcal{G}_{\hat{A}\hat{B}}(X)) = \begin{pmatrix} G_{ij} & 0 & 0 \\ 0 & h_{\alpha\beta} & 0 \\ 0 & 0 & \varepsilon_s \mathrm{e}^{-2U^s(X)} \delta_{ss'} \end{pmatrix}. \tag{3.14}$$

The minisuperspace metric (3.12) may be also written in the form

$$\mathcal{G} = \bar{G} + \sum_{s \in S} \varepsilon_s \mathrm{e}^{-2U^s(x)} d\Phi^s \otimes d\Phi^s, \tag{3.15}$$

where $x = (x^A) = (\phi^i, \varphi^\alpha)$,

$$\bar{G} = \bar{G}_{AB} dx^A \otimes dx^B = G_{ij} d\phi^i \otimes d\phi^j + h_{\alpha\beta} d\varphi^\alpha \otimes d\varphi^\beta, \tag{3.16}$$

$$(\bar{G}_{AB}) = \begin{pmatrix} G_{ij} & 0 \\ 0 & h_{\alpha\beta} \end{pmatrix}, \tag{3.17}$$

$U^s(x) = U^s_A x^A$ is defined in (3.6) and

$$(U^s_A) = (d_i \delta_{iI_s}, -\chi_s \lambda_{a_s \alpha}). \tag{3.18}$$

Here

$$\delta_{iI} \equiv \sum_{j \in I} \delta_{ij} = \begin{matrix} 1, \ i \in I \\ 0, \ i \notin I \end{matrix} \tag{3.19}$$

is an indicator of i belonging to I. The potential (3.3) reads

$$V = (-w\Lambda) \mathrm{e}^{2U^\Lambda(x)} + \sum_{j=0}^{n} \frac{w}{2} \xi_j d_j \mathrm{e}^{2U^j(x)}, \tag{3.20}$$

where

$$U^j(x) = U^j_A x^A = -\phi^j + \gamma_0(\phi), \tag{3.21}$$

$$U^\Lambda(x) = U^\Lambda_A x^A = \gamma_0(\phi), \tag{3.22}$$

$$(U^j_A) = (-\delta^j_i + d_i, 0), \tag{3.23}$$

$$(U^\Lambda_A) = (d_i, 0). \tag{3.24}$$

The integrability of the Lagrange system (3.11) depends upon the scalar products of co-vectors U^Λ, U^j, U^s corresponding to $\bar{G}$:

$$(U, U') = \bar{G}^{AB} U_A U'_B, \tag{3.25}$$

where

$$(\bar{G}^{AB}) = \begin{pmatrix} G^{ij} & 0 \\ 0 & h^{\alpha\beta} \end{pmatrix} \tag{3.26}$$

is matrix inverse to (3.17). Here (as in [52])

$$G^{ij} = \frac{\delta^{ij}}{d_i} + \frac{1}{2 - D}, \tag{3.27}$$

$i, j = 0, \dots, n$. These products have the following form

$$(U^i, U^j) = \frac{\delta_{ij}}{d_j} - 1, \tag{3.28}$$

$$(U^\Lambda, U^\Lambda) = -\frac{D-1}{D-2}, \tag{3.29}$$

$$(U^s, U^{s'}) = q(I_s, I_{s'}) + \chi_s \chi_{s'} \lambda_{a_s} \cdot \lambda_{a_{s'}}, \tag{3.30}$$

$$(U^s, U^i) = -\delta_{iI_s}, \tag{3.31}$$

where $s = (a_s, v_s, I_s)$, $s' = (a_{s'}, v_{s'}, I_{s'}) \in S$,

$$q(I, J) \equiv d(I \cap J) + \frac{d(I)d(J)}{2 - D}, \tag{3.32}$$

$$\lambda_a \cdot \lambda_b \equiv \lambda_{a\alpha} \lambda_{b\beta} h^{\alpha\beta}. \tag{3.33}$$

Relations (3.28)-(3.29) were found in [70] and (3.30) in [143].

4 Cosmological and Spherically Symmetric Solutions

Here we put the following restrictions on the parameters of the model

$$(\mathbf{i}) \qquad \Lambda = 0, \tag{4.1}$$

i.e. the cosmological constant is zero,

$$(\mathbf{ii}) \qquad \xi_0 \neq 0, \quad \xi_1 = \ldots = \xi_n = 0, \tag{4.2}$$

i.e. one space is curved and others are Ricci-flat,

$$(\mathbf{iii}) \qquad 0 \notin I_s, \quad \forall s = (a_s, v_s, I_s) \in S, \tag{4.3}$$

i.e. all “brane” manifolds M_{I_s} (see (2.19)) do not contain M_0.

We also impose a block-orthogonality restriction on the set of vectors $(U^s, s \in S)$. Let

$$S = S_1 \sqcup \ldots \sqcup S_k, \tag{4.4}$$

$S_i \neq \emptyset$, $i = 1, \ldots, k$, and

$$(\mathbf{iv}) \ (U^s, U^{s'}) = d(I_s \cap I_{s'}) + \frac{d(I_s)d(I_{s'})}{2 - D} + \chi_s \chi_{s'} \lambda_{a_s \alpha} \lambda_{a_{s'} \beta} h^{\alpha\beta} = 0, \tag{4.5}$$

for all $s = (a_s, v_s, I_s) \in S_i$, $s' = (a_{s'}, v_{s'}, I_{s'}) \in S_j$, $i \neq j$; $i, j = 1, \ldots, k$. Relation (4.4) means that the set S is a union of k non-intersecting (non-empty) subsets $S_1, \ldots, S_k$. According to (4.5) the set of vectors $(U^s, s \in S)$ has a block-orthogonal structure with respect to the scalar product (3.25), i.e. it splits into k mutually orthogonal blocks $(U^s, s \in S_i)$, $i = 1, \ldots, k$.

From **(i)**, **(ii)** we get for the potential (3.20)

$$V = \frac{1}{2} w \xi_0 d_0 \mathrm{e}^{2U^0(x)}, \tag{4.6}$$

where

$$(U^0, U^0) = \frac{1}{d_0} - 1 < 0 \tag{4.7}$$

(see (3.28)).

From **(iii)** and (3.31) we get

$$(U^0, U^s) = 0 \tag{4.8}$$

for all $s \in S$. Thus, the set of co-vectors U^0, U^s, $s \in S$ (belonging to dual space $(\mathbb{R}^{n+1+l})^* \simeq \mathbb{R}^{n+1+l}$) has also a block-orthogonal structure with respect to the scalar product (3.25).

Here we fix the time gauge as follows

$$\gamma = \gamma_0, \quad \mathcal{N} = 1, \tag{4.9}$$

i.e the harmonic time gauge is used. Then we obtain the Lagrange system with the Lagrangian

$$L = \frac{\mu_*}{2} \mathcal{G}_{\hat{A}\hat{B}}(X) \dot{X}^{\hat{A}} \dot{X}^{\hat{B}} - \mu_* V \tag{4.10}$$

and the energy constraint

$$E = \frac{\mu_*}{2} \mathcal{G}_{\hat{A}\hat{B}}(X) \dot{X}^{\hat{A}} \dot{X}^{\hat{B}} + \mu_* V = 0. \tag{4.11}$$

Here we will integrate the Lagrange equations corresponding to the Lagrangian (4.10) with the energy-constraint (4.11) and hence we will find classical exact solutions when the restrictions (8.2), (8.3) from Appendix 1 are imposed.

The problem of integrability may be simplified if we integrate the Maxwell equations (for $s \in S_e$) and Bianchi identities (for $s \in S_m$):

$$\frac{d}{du} \left(\exp(-2U^s) \dot{\Phi}^s \right) = 0 \Longleftrightarrow \dot{\Phi}^s = Q_s \exp(2U^s), \tag{4.12}$$

where Q_s are constants, $s \in S$.

Let

$$Q_s \neq 0, \tag{4.13}$$

for all $s \in S$.

For fixed $Q = (Q_s, s \in S)$ the Lagrange equations for the Lagrangian (4.10) corresponding to $(x^A) = (\phi^i, \varphi^\alpha)$, when equations (4.12) are substituted are equivalent to the Lagrange equations for the Lagrangian

$$L_Q = \frac{1}{2} \bar{G}_{AB} \dot{x}^A \dot{x}^B - V_Q, \tag{4.14}$$

where

$$V_Q = V + \frac{1}{2} \sum_{s \in S} \varepsilon_s Q_s^2 \exp[2U^s(x)], \tag{4.15}$$

$(\bar{G}_{AB})$ and V are defined in (3.17) and (4.6) respectively. The zero-energy constraint (4.11) reads

$$E_Q = \frac{1}{2} \bar{G}_{AB} \dot{x}^A \dot{x}^B + V_Q = 0. \tag{4.16}$$

When the conditions **(i)–(iv)** are satisfied exact solutions to Lagrange equations corresponding to (4.14) with the potential (4.15) and V from (4.6) could be readily obtained using the relations from Appendix 2.

The solutions read:

$$x^A(u) = -\frac{U^{0A}}{(U^0, U^0)} \ln |f_0(u)| - \sum_{s \in S} \eta_s \nu_s^2 U^{sA} \ln |f_s(u)| + c^A u + \bar{c}^A. \quad (4.17)$$

Functions f_0 and f_s in (4.17) are the following:

$$f_0(u) = |\xi_0(d_0 - 1)|^{1/2} s(u - u_0, w\xi_0, C_0) = \quad (4.18)$$

$$\left| \frac{\xi_0(d_0 - 1)}{C_0} \right|^{1/2} \mathrm{sh}(\sqrt{C_0}(u - u_0)), \ C_0 > 0, \ \xi_0 w > 0; \quad (4.19)$$

$$\left| \frac{\xi_0(d_0 - 1)}{C_0} \right|^{1/2} \sin(\sqrt{|C_0|}(u - u_0)), \ C_0 < 0, \ \xi_0 w > 0; \quad (4.20)$$

$$\left| \frac{\xi_0(d_0 - 1)}{C_0} \right|^{1/2} \mathrm{ch}(\sqrt{C_0}(u - u_0)), \ C_0 > 0, \ \xi_0 w < 0; \quad (4.21)$$

$$|\xi_0(d_0 - 1)|^{1/2} (u - u_0), \ C_0 = 0, \ \xi_0 w > 0, \quad (4.22)$$

and

$$f_s(u) = \frac{|Q_s|}{|\nu_s|} s(u - u_s, -\eta_s \varepsilon_s, C_s) = \quad (4.23)$$

$$\frac{|Q_s|}{|\nu_s||C_s|^{1/2}} \mathrm{sh}(\sqrt{C_s}(u - u_s)), \ C_s > 0, \ \eta_s \varepsilon_s < 0; \quad (4.24)$$

$$\frac{|Q_s|}{|\nu_s||C_s|^{1/2}} \sin(\sqrt{|C_s|}(u - u_s)), \ C_s < 0, \ \eta_s \varepsilon_s < 0; \quad (4.25)$$

$$\frac{|Q_s|}{|\nu_s||C_s|^{1/2}} \mathrm{ch}(\sqrt{C_s}(u - u_s)), \ C_s > 0, \ \eta_s \varepsilon_s > 0; \quad (4.26)$$

$$\frac{|Q^s|}{|\nu_s|} (u - u_s), \ C_s = 0, \ \eta_s \varepsilon_s < 0, \quad (4.27)$$

where C_0, C_s, u_0, u_s are constants, $s \in S$. The function $s(u, \xi, C)$ is defined in Appendix 2.

The parameters $\eta_s = \pm 1$, $\nu_s \neq 0$, $s \in S$, satisfy the relations

$$\sum_{s' \in S} (U^s, U^{s'}) \eta_{s'} \nu_{s'}^2 = 1, \quad (4.28)$$

for all $s \in S$, with scalar products $(U^s, U^{s'})$ defined in (3.30).

The constants C_s, u_s are coinciding inside blocks:

$$u_s = u_{s'}, \qquad C_s = C_{s'}, \quad (4.29)$$

$s, s' \in S_i$, $i = 1, \dots, k$ (see relation (9.13) from Appendix 2). The ratios $\varepsilon_s Q_s^2/(\eta_s \nu_s^2)$ are also coinsiding inside blocks, or, equivalently,

$$\varepsilon_s \eta_s = \varepsilon_{s'} \eta_{s'} \tag{4.30}$$

$$\frac{Q_s^2}{\nu_s^2} = \frac{Q_{s'}^2}{\nu_{s'}^2}, \tag{4.31}$$

$s, s' \in S_i$, $i = 1, \dots, k$. Here we used the relations (4.7), (4.8).

The contravariant components $U^{rA} = \bar{G}^{AB} U_B^r$ are [166]

$$U^{0i} = -\frac{\delta_0^i}{d_0}, \quad U^{0\alpha} = 0, \tag{4.32}$$

$$U^{si} = G^{ij} U_j^s = \delta_{iI_s} - \frac{d(I_s)}{D-2}, \quad U^{s\alpha} = -\chi_s \lambda_{a_s}^{\alpha}. \tag{4.33}$$

Using (4.17), (4.7), (4.33) and (4.32) we obtain

$$\phi^i = \frac{\delta_0^i}{1-d_0} \ln|f_0| - \sum_{s \in S} \eta_s \nu_s^2 \left(\delta_{iI_s} - \frac{d(I_s)}{D-2} \right) \ln|f_s| + c^i u + \bar{c}^i, \tag{4.34}$$

and

$$\varphi^{\alpha} = \sum_{s \in S} \eta_s \nu_s^2 \chi_s \lambda_{a_s}^{\alpha} \ln|f_s| + c^{\alpha} u + \bar{c}^{\alpha}, \tag{4.35}$$

$\alpha = 1, \dots, l$.

Vectors $c = (c^A)$ and $\bar{c} = (\bar{c}^A)$ satisfy the linear constraint relations (see (9.20) in Appendix 2)

$$U^0(c) = U_A^0 c^A = -c^0 + \sum_{j=0}^{n} d_j c^j = 0, \tag{4.36}$$

$$U^0(\bar{c}) = U_A^0 \bar{c}^A = -\bar{c}^0 + \sum_{j=0}^{n} d_j \bar{c}^j = 0, \tag{4.37}$$

$$U^s(c) = U_A^s c^A = \sum_{i \in I_s} d_i c^i - \chi_s \lambda_{a_s \alpha} c^{\alpha} = 0, \tag{4.38}$$

$$U^s(\bar{c}) = U_A^s \bar{c}^A = \sum_{i \in I_s} d_i \bar{c}^i - \chi_s \lambda_{a_s \alpha} \bar{c}^{\alpha} = 0, \tag{4.39}$$

$s \in S$. The (3.4) reads

$$\gamma_0(\phi) = \frac{d_0}{1-d_0} \ln|f_0| + \sum_{s \in S} \frac{d(I_s)}{D-2} \eta_s \nu_s^2 \ln|f_s| + c^0 u + \bar{c}^0. \tag{4.40}$$

The zero-energy constraint reads (see Appendix 2)

$$E = E_0 + \sum_{s \in S} E_s + \frac{1}{2} \bar{G}_{AB} c^A c^B = 0, \tag{4.41}$$

where $E_0 = C_0(U^0, U^0)^{-1}/2$, $E_s = C_s(\eta_s \nu_s^2)/2$. Using relations (3.10), (3.17), (4.7) and (4.36) we rewrite (4.41) as

$$C_0 \frac{d_0}{d_0 - 1} = \sum_{s \in S} C_s \nu_s^2 \eta_s + h_{\alpha\beta} c^\alpha c^\beta + \sum_{i=1}^{n} d_i (c^i)^2 + \frac{1}{d_0 - 1} \left(\sum_{i=1}^{n} d_i c^i \right)^2 . \quad (4.42)$$

From relation

$$\exp(2U^s) = f_s^{-2}, \quad (4.43)$$

following from (4.5), (4.8), (4.17), (4.38) and (4.39) we get for electric-type forms (2.22)

$$\mathcal{F}^s = Q_s f_s^{-2} du \wedge \tau(I_s), \quad (4.44)$$

$s \in S_e$, and for magnetic-type forms (2.23)

$$\mathcal{F}^s = \mathrm{e}^{-2\lambda_a(\varphi)} * \left[Q_s f_s^{-2} du \wedge \tau(I_s) \right] = \bar{Q}_s \tau(\bar{I}_s), \quad (4.45)$$

$s \in S_m$, where $\bar{Q}_s = Q_s \varepsilon(I_s) \mu(I_s) w$ and $\mu(I) = \pm 1$ is defined by the relation $\mu(I) du \wedge \tau(I_0) = \tau(\bar{I}) \wedge du \wedge \tau(I)$. The relation (4.45) follows from the formula (5.26) from [143] (for $\gamma = \gamma_0$).

Relations for the metric follows from (4.34) and (4.40)

$$g = \left(\prod_{s \in S} [f_s^2(u)]^{\eta_s d(I_s) \nu_s^2/(D-2)} \right) \Big\{ [f_0^2(u)]^{d_0/(1-d_0)} \mathrm{e}^{2c^0 u + 2\bar{c}^0} \quad (4.46)$$
$$\times [w du \otimes du + f_0^2(u) g^0] + \sum_{i=1}^{n} \Big(\prod_{s \in S} [f_s^2(u)]^{-\eta_s \nu_s^2 \delta_{i I_s}} \Big) \mathrm{e}^{2c^i u + 2\bar{c}^i} g^i \Big\}.$$

Thus, here we obtained the "block-orthogonal" generalization of the solution from [166]. This solution describes the evolution of $n + 1$ spaces $(M_0, g_0), \ldots, (M_n, g_n)$, where (M_0, g_0) is an Einstein space of non-zero curvature, and (M_i, g^i) are "internal" Ricci-flat spaces, $i = 1, \ldots, n$; in the presence of several scalar fields and forms. The solution is presented by relations (4.35), (4.44)-(4.46) with the functions f_0, f_s defined in (4.18)–(4.27) and the relations on the parameters of solutions c^A, $\bar{c}^A$ $(A = i, \alpha)$, C_0, C_s, u_s, Q_s , η_s, ν_s $(s \in S)$ imposed in (4.28)–(4.31), (4.36)–(4.39), (4.42), respectively.

This solution describes a set of charged (by forms) overlapping p-branes ($p_s = d(I_s) - 1$, $s \in S$) "living" on submanifolds (isomorphic to) M_{I_s} (2.19), where the sets I_s do not contain 0, i.e. all p-branes live in "internal" Ricci-flat spaces.

The solution is valid if the dimensions of p-branes and dilatonic coupling vector satisfy the relations (4.5). In "orthogonal" non-composite case these solutions were considered in [154,153] (electric case) and [151] (electro-magnetic case). For $n = 1$ see also [156,130]. In block-orthogonal (non-composite) case a special class of solutions with ν_s^2 coinciding inside blocks was considered earlier in [167].

5 Spherically Symmetric and Black Hole Solutions

Here we consider the spherically symmetric case

$$w = 1, \quad M_0 = S^{d_0}, \quad g^0 = d\Omega^2_{d_0}, \tag{5.1}$$

where $d\Omega^2_{d_0}$ is the canonical metric on a unit sphere S^{d_0}, $d \geq 2$. We also assume that

$$M_1 = \mathbb{R}, \quad g^1 = -dt \otimes dt \tag{5.2}$$

(here M_1 is a time manifold) and

$$1 \in I_s, \quad \forall s \in S, \tag{5.3}$$

i. e. all p-branes have a common time direction t.

For integration constants we put $\bar{c}^A = 0$,

$$c^A = \bar{\mu}(\bar{b}^A - b^A), \tag{5.4}$$

$$\bar{b}^A = \bar{\mu} \sum_{r \in \bar{S}} \eta_r \nu_r^2 U^{rA} - \bar{\mu}\delta_1^A, \tag{5.5}$$

$$C_0 = \bar{\mu}^2, \tag{5.6}$$

$$C_s = \bar{\mu}^2 b_s^2, \qquad b_s > 0, \tag{5.7}$$

where $\bar{\mu} > 0$, $\bar{S} = \{0\} \cup S$ and $\eta_0 \nu_0^2 = (U^0, U^0)^{-1}$.

The only essential restrictions imposed are the inequalities $C_0, C_s > 0$ that cut a subclass in the class of solutions from Section 4. This subclass contains non-extremal black hole solutions and its "Kasner-like" (non-black-hole) deformations. For extremal black hole solutions one should consider the special case $C_0 = C_s = 0$. (For extremal black hole solutions and its multicenter generalizations see [168].)

Due to (4.29) the parameters b_s, $s \in S$, are coinciding inside blocks:

$$b_s = b_{s'}, \tag{5.8}$$

$s, s' \in S_i$, $i = 1, \ldots, k$.

It may be verified that the restrictions (4.36) and (4.38) are satisfied identically if and only if

$$U^0(b) = U^0_A b^A = -b^0 + \sum_{j=0}^{n} d_j b^j = 1, \tag{5.9}$$

$$U^s(b) = U^s_A b^A = \sum_{i \in I_s} d_i b^i - \chi_s \lambda_{a_s \alpha} b^\alpha = 1, \tag{5.10}$$

$s \in S$. This follows from identities $U^0(\bar{b}) = 1$ and $U^s(\bar{b}) = 1$, $s \in S$.

Relation (4.42) reads

$$\sum_{s\in S}\eta_s\nu_s^2(b_s^2-1)+h_{\alpha\beta}b^\alpha b^\beta+\sum_{i=1}^n d_i(b^i)^2+\frac{1}{d_0-1}\left(\sum_{i=1}^n d_i b^i\right)^2=\frac{d_0}{d_0-1}, \tag{5.11}$$

where the relation (equivalent to (5.9))

$$b^0=\frac{1}{1-d_0}\left[\sum_{j=1}^n d_j b^j-1\right], \tag{5.12}$$

is used.

Now we rewrite a solution (under restrictions imposed) in a so-called "Kasner-like" form that is more suitable for analysing the behaviour at large distances and for singling out the black hole solutions. For this reason we introduce a new radial variable $R=R(u)$ by relations

$$\exp(-2\bar{\mu}u)=1-\frac{2\mu}{R^{\bar{d}}},\quad \mu=\bar{\mu}/\bar{d}>0,\quad \bar{d}=d_0-1, \tag{5.13}$$

$u>0$, $R^{\bar{d}}>2\mu$. For the function

$$f_s(u)=\frac{|Q_s|}{2\bar{\mu}b_s|\nu_s|}[\exp(\bar{\mu}b_s(u-u_s))+\eta_s\varepsilon_s\exp(-\bar{\mu}b_s(u-u_s))] \tag{5.14}$$

we put the restriction $f_s(0)=1$, or, equivalently,

$$\exp(-\bar{\mu}b_s u_s)+\eta_s\varepsilon_s\exp(\bar{\mu}b_s u_s)=\frac{2\bar{\mu}b_s|\nu_s|}{|Q_s|}. \tag{5.15}$$

This restriction guarantees the asymptotical flatness of the $(2+d_0)$-dimensional section of the metric in the limit $R\to+\infty$ (or, when, $u\to+0$)). It follows from (5.15) that $u_s<0$ for $\eta_s\varepsilon_s=-1$. In any case $f_s(u)>0$ for $u\geq 0$.

Then, solutions for the metric and scalar fields (see (4.35), (4.46)) may be written as follows

$$g=\Bigl(\prod_{s\in S}\bar{H}_s^{2\eta_s d(I_s)\nu_s^2/(D-2)}\Bigr)\Bigl\{F^{b^0-1}dR\otimes dR+R^2F^{b^0}d\Omega^2_{d_0} \tag{5.16}$$

$$-\Bigl(\prod_{s\in S}\bar{H}_s^{-2\eta_s\nu_s^2}\Bigr)F^{b^1}dt\otimes dt+\sum_{i=2}^n\Bigl(\prod_{s\in S}\bar{H}_s^{-2\eta_s\nu_s^2\delta_{iI_s}}\Bigr)F^{b^i}g^i\Bigr\},$$

$$\varphi^\alpha=\sum_{s\in S}\eta_s\nu_s^2\chi_s\lambda^\alpha_{a_s}\ln\bar{H}_s+\frac{1}{2}b^\alpha\ln F, \tag{5.17}$$

where

$$F=1-\frac{2\mu}{R^{\bar{d}}}, \tag{5.18}$$

$$\bar{H}_s=\hat{H}_sF^{(1-b_s)/2}, \tag{5.19}$$

$$\hat{H}_s=1+\hat{P}_s\frac{(1-F^{b_s})}{2\mu b_s}, \tag{5.20}$$

$$\hat{P}_s = -\varepsilon_s \eta_s P_s, \tag{5.21}$$

$$P_s = \frac{|Q_s|}{\bar{d}|\nu_s|} \exp(\mu u_s) > 0, \tag{5.22}$$

$s \in S$. Due to (4.29)-(4.31) parameters P_s and $\hat{P}_s$ are coinciding inside blocks:

$$P_s = P_{s'}, \qquad \hat{P}_s = \hat{P}_{s'}, \tag{5.23}$$

$s, s' \in S_i$, $i = 1, \ldots, k$. Parameters b_s are also coinciding inside blocks, see (5.8). Parameters b_s, b^i, b^α obey the relations (5.10)-(5.12).

The fields of forms are given by (2.22), (2.23) with

$$\Phi^s = \frac{\nu_s}{H'_s}, \tag{5.24}$$

$$H'_s = \left[1 - P'_s \hat{H}_s^{-1} \frac{(1 - F^{b_s})}{2\mu b_s}\right]^{-1}, \tag{5.25}$$

$$P'_s = -\frac{Q_s}{\nu_s \bar{d}}. \tag{5.26}$$

$s \in S$. It follows from (5.15), (5.20), (5.21) and (5.26) that

$$(P'_s)^2 = P_s(\hat{P}_s + 2b_s\mu) = -\varepsilon_s \eta_s \hat{P}_s(\hat{P}_s + 2b_s\mu), \tag{5.27}$$

$s \in S$. This relation is self-consistent, i.e. its left- and right-hand sides have the same sign, since due to (5.15) and (5.22)

$$P_s < 2\mu b_s \tag{5.28}$$

for $\varepsilon_s \eta_s = +1$ and hence

$$\hat{P}_s > -2b_s\mu, \tag{5.29}$$

for all $s \in S$.

5.1 Black Hole Solutions

Here we show that the black hole solution from [168] may be obtained from our spherically-symmetric solutions (5.16)-(5.27) when

$$b^1 = b_s = 1, \qquad b^i = b^\alpha = 0, \tag{5.30}$$

$s \in S$, $i = 0, 2, \ldots, n$, $\alpha = 1, \ldots, l$.

Under relations (5.30) imposed the metric and scalar fields (5.16) and (5.17) read

$$g = \Big(\prod_{s \in S} \hat{H}_s^{2\eta_s d(I_s)\nu_s^2/(D-2)}\Big)\Big\{\frac{dR \otimes dR}{1 - 2\mu/R^{\bar{d}}} + R^2 d\Omega^2_{d_0} \tag{5.31}$$
$$-\Big(\prod_{s \in S} \hat{H}_s^{-2\eta_s \nu_s^2}\Big)\left(1 - \frac{2\mu}{R^{\bar{d}}}\right) dt \otimes dt + \sum_{i=2}^{n} \Big(\prod_{s \in S} \hat{H}_s^{-2\eta_s \nu_s^2 \delta_{iI_s}}\Big) g^i\Big\},$$

$$\varphi^\alpha = \sum_{s \in S} \eta_s \nu_s^2 \chi_s \lambda^\alpha_{a_s} \ln \hat{H}_s, \tag{5.32}$$

where $\mu > 0$, $R^{\bar{d}} > 2\mu$ and

$$\hat{H}_s = 1 + \frac{\hat{P}_s}{R^{\bar{d}}}, \quad \hat{P}_s > -2\mu, \tag{5.33}$$

$\hat{P}_s \neq 0$, $s \in S$. Parameters $\hat{P}_s$ are coinciding inside blocks (see (5.23)).

The fields of forms are given by (2.22), (2.23) with

$$\Phi^s = \frac{\nu_s}{H'_s}, \tag{5.34}$$

$$H'_s = \left(1 - \frac{P'_s}{\hat{H}_s R^{\bar{d}}}\right)^{-1} = 1 + \frac{P'_s}{R^{\bar{d}} + \hat{P}_s - P'_s}, \tag{5.35}$$

$s \in S$. Here

$$(P'_s)^2 = -\varepsilon_s \eta_s \hat{P}_s (\hat{P}_s + 2\mu), \tag{5.36}$$

and

$$\varepsilon_s \eta_s \hat{P}_s < 0, \tag{5.37}$$

$s \in S$. Parameters ν_s satisfy relations (4.28).

The solution obtained describes non-extremal charged p-brane black holes with block-orthogonal intersection rules. The exteriour horizon corresponds to $R^{\bar{d}} \to 2\mu$.

Let

$$\varepsilon_s \eta_s = -1, \tag{5.38}$$

$s \in S$. This restriction is satisfied in orthogonal case, when pseudo-Euclidean p-branes in a space-time of pseudo-Euclidean signature are considered (in this case all $\varepsilon(I_s) = -1$, $\varepsilon[g] = -1$), all $\theta_s = +1$ in the action (2.1) and $\eta_s = \mathrm{sign}(U^s, U^s) = +1$).

Under restrictions (5.38) imposed our solutions agree with the solutions with orthogonal intersection rules from Refs. [131], [149], [150] ($d_1 = \ldots = d_n = 1$, $\eta_s = +1$), [151] ($\eta_s = +1$, non-composite case) and block-orthogonal ones from [167] (for ν_s coinciding inside blocks).

Hawking Temperature. The Hawking temperature corresponding to the solution is (see also [151,150])

$$T_H(\mu) = \frac{\bar{d}}{4\pi (2\mu)^{1/\bar{d}}} \prod_{s \in S} \left(\frac{2\mu}{2\mu + \hat{P}_s}\right)^{\eta_s \nu_s^2}. \tag{5.39}$$

For fixed $\hat{P}_s > 0$ ($\varepsilon_s \eta_s = -1$) and $\mu \to +0$ we get $T_H(\mu) \to 0$ for the extremal black hole configurations [168] satisfying

$$\xi = \sum_{s \in S} \eta_s \nu_s^2 - \bar{d}^{-1} > 0. \tag{5.40}$$

6 Post-Newtonian Approximation

Let $d_0 = 2$. Here we consider the 4-dimensional section of the metric (5.16)

$$g^{(4)} = U\Big\{F^{b^0-1} dR \otimes dR + F^{b^0} R^2 d\hat{\Omega}_2^2 - U_1 F^{b^1} dt \otimes dt\Big\}, \tag{6.1}$$

where $F = 1 - (2\mu/R)$, and

$$U = \prod_{s \in S} \bar{H}_s^{2\eta_s d(I_s)\nu_s^2/(D-2)}, \tag{6.2}$$

$$U_1 = \prod_{s \in S} \bar{H}_s^{-2\eta_s \nu_s^2}, \tag{6.3}$$

$$U_i = \prod_{s \in S} \bar{H}_s^{-2\eta_s \nu_s^2 \delta_{iI_s}}, \qquad i > 1, \tag{6.4}$$

$R > 2\mu$.

We may suppose that some real astrophysical objects (e.g. stars) are described by the 4-dimensional "physical" metric (6.1), i.e. they are "traces" of extended multidimensional objects (charged p-branes).

Introducing a new radial variable ρ by the relation

$$R = \rho \left(1 + \frac{\mu}{2\rho}\right)^2, \tag{6.5}$$

($\rho > \mu/2$), we rewrite the metric (6.1) in a 3-dimensional conformally-flat form

$$g^{(4)} = U\Big\{-U_1 F^{b^1} dt \otimes dt + F^{b^0}\left(1 + \frac{\mu}{2\rho}\right)^4 \delta_{ij} dx^i \otimes dx^j\Big\}, \tag{6.6}$$

$$F = \left(1 - \frac{\mu}{2\rho}\right)^2 \left(1 + \frac{\mu}{2\rho}\right)^{-2} \tag{6.7}$$

where $\rho^2 = |x|^2 = \delta_{ij} x^i x^j$ $(i, j = 1, 2, 3)$.

For possible physical applications we should calculate the post-Newtonian parameters β and γ (Eddington parameters) using the following relations (see, for example, [170] and references therein)

$$g_{00}^{(4)} = -(1 - 2V + 2\beta V^2) + O(V^3), \tag{6.8}$$

$$g_{ij}^{(4)} = \delta_{ij}(1 + 2\gamma V) + O(V^2), \tag{6.9}$$

$i, j = 1, 2, 3$, where

$$V = \frac{GM}{\rho} \tag{6.10}$$

is the Newton's potential, G is the gravitational constant, M is the gravitational mass. From (6.6)-(6.10) we get

$$GM = \mu b^1 + \sum_{s \in S} \eta_s \nu_s^2 [\hat{P}_s + (b_s - 1)\mu] \left(1 - \frac{d(I_s)}{D-2}\right) \tag{6.11}$$

and for $GM \neq 0$

$$\beta - 1 = \frac{1}{2(GM)^2} \sum_{s \in S} \eta_s \nu_s^2 \hat{P}_s (\hat{P}_s + 2 b_s \mu) \left(1 - \frac{d(I_s)}{D-2}\right) \tag{6.12}$$

$$\gamma - 1 = -\frac{1}{GM} \left[\mu(b^0 + b^1 - 1) + \sum_{s \in S} \eta_s \nu_s^2 [\hat{P}_s + (b_s - 1)\mu] \left(1 - 2\frac{d(I_s)}{D-2}\right) \right] \tag{6.13}$$

It follows from (5.27), (6.12) and the inequalities $d(I_s) < D - 2$ (for all $s \in S$) that the following inequalities take place

$$\beta > 1, \text{ if all } \varepsilon_s = -1, \tag{6.14}$$

$$\beta < 1, \text{ if all } \varepsilon_s = +1. \tag{6.15}$$

There exists a large variety of configurations with $\beta = 1$ when the relation $\varepsilon_s =$ const is broken.

There exist also non-trivial p-brane configurations with $\gamma = 1$.

Proposition. Let the set of p-branes consist of several pairs of electric and magnetic branes. Let any such pair $(s, \bar{s} \in S)$ correspond to the same colour index, i.e. $a_s = a_{\bar{s}}$, and $\hat{P}_s = \hat{P}_{\bar{s}}$, $b_s = b_{\bar{s}}$, $\eta_s \nu_s^2 = \eta_{\bar{s}} \nu_{\bar{s}}^2$. Then for $b^0 + b^1 = 1$ we get

$$\gamma = 1. \tag{6.16}$$

The Proposition can be readily proved using the relation $d(I_s) + d(I_{\bar{s}}) = D - 2$, following from (2.28) and (2.29).

Observational Restrictions. The observations in the solar system give the tight constraints on the Eddington parameters [170]

$$\gamma = 1.000 \pm 0.002 \tag{6.17}$$

$$\beta = 0.9998 \pm 0.0006. \tag{6.18}$$

The first restriction is a result of the Viking time-delay experiment [171]. The second restriction follows from (6.17) and the analysis of the laser ranging data to the Moon. In this case a high precision test based on the calculation of the combination $(4\beta - \gamma - 3)$ appearing in the Nordtvedt effect [173] is used [172]. We note, that as it was pointed in [170] the "classic" tests of general relativity, i.e. the Mercury-perihelion and light deflection tests, are somewhat outdated.

For small enough $\hat{p}_s = \hat{P}_s/GM$, $b_s - 1$, $b^1 - 1$, b^i $(i > 1)$ of the same order we get $GM \sim \mu$ and hence

$$\beta - 1 \sim \sum_{s \in S} \eta_s \nu_s^2 \hat{p}_s \left(1 - \frac{d(I_s)}{D-2}\right) \tag{6.19}$$

$$\gamma - 1 \sim -b^0 - b^1 + 1 - \sum_{s \in S} \eta_s \nu_s^2 [\hat{p}_s + (b_s - 1)] \left(1 - 2\frac{d(I_s)}{D-2}\right), \tag{6.20}$$

i.e. $\beta - 1$ and $\gamma - 1$ are of the same order. Thus for small enough $\hat{p}_s$, $b_s - 1$, $b^1 - 1$, b^i $(i > 1)$ it is possible to fit the "solar system" restrictions (6.17) and (6.18).

There exists also another possibility to satisfy these restrictions.

One Brane Case. Let us consider a special case of one p-brane. In this case we have

$$\eta_s \nu_s^{-2} = d(I_s) \left(1 - \frac{d(I_s)}{D-2}\right) + \lambda^2. \tag{6.21}$$

Relations (6.12), (6.13) and (6.21) imply that for large enough values of (dilatonic coupling constant squared) λ^2 and $b^0 + b^1 = 1$ it is possible to perform the "fine tuning" the parameters (β, γ) near the point (1, 1) even if the parameters $\hat{P}_s$ are big.

7 Conclusions

In this paper we obtained exact solutions to Einstein equations for the multi-dimensional cosmological model describing the evolution of n Ricci-flat spaces and one Einstein space M_0 of non-zero curvature in the presence of composite electro-magnetic p-branes. The solutions were obtained in the block-orthogonal case (4.5), when p-branes do not "live" in M_0. We also considered the spherically-symmetric solutions containing non-extremal p-brane black holes [167,168]. The relations for post-Newtonian parameters β and γ are obtained.

Acknowledgments. This work was supported in part by DFG grant 436 RUS 113/236/O(R) and by the Russian Ministry for Science and Technology, Russian Fund for Basic Research, project N 98-02-16414 and project SEE.

Appendix 1: Restrictions on *p*-Brane Configurations

Restrictions on $\Omega_{a,v}$ [166]. Let

$$w_1 \equiv \{i \mid i \in \{0, \dots, n\},\ d_i = 1\}. \tag{8.1}$$

The set w_1 describes all 1-dimensional manifolds among M_i $(i \geq 0)$. We impose the following restrictions on the sets $\Omega_{a,v}$ (2.24):

$$W_{ij}(\Omega_{a,v}) = \emptyset, \tag{8.2}$$

$a \in \Delta$; $v = e, m$; $i, j \in w_1$, $i < j$ and

$$W_j^{(1)}(\Omega_{a,m}, \Omega_{a,e}) = \emptyset, \tag{8.3}$$

$a \in \Delta$; $j \in w_1$. Here

$$W_{ij}(\Omega_*) \equiv \{(I, J)|I, J \in \Omega_*,\ I = \{i\} \sqcup (I \cap J),\ J = \{j\} \sqcup (I \cap J)\}, \tag{8.4}$$

$i, j \in w_1$, $i \neq j$, $\Omega_* \subset \Omega_0$ and

$$W_j^{(1)}(\Omega_{a,m}, \Omega_{a,e}) \equiv \{(I, J) \in \Omega_{a,m} \times \Omega_{a,e} | \bar{I} = \{j\} \sqcup J\}, \tag{8.5}$$

$j \in w_1$. In (8.5)

$$\bar{I} \equiv I_0 \setminus I \tag{8.6}$$

is "dual" set, $(I_0 = \{0, 1, \dots, n\})$.

The restrictions (8.2) and (8.3) are trivially satisfied when $n_1 \leq 1$ and $n_1 = 0$ respectively, where $n_1 = |w_1|$ is the number of 1-dimensional manifolds among M_i. They are also satisfied in the non-composite case when all $|\Omega_{a,v}| = 1$. For $n_1 \geq 2$ and $n_1 \geq 1$, respectively, these restrictions forbid certain pairs of two p-branes, corresponding to the same form $F^a, a \in \Delta$:

Appendix 2: Solutions with Block-Orthogonal Set of Vectors

Let

$$L = \frac{1}{2} < \dot{x}, \dot{x} > - \sum_{s \in S} A_s \exp(2 < b_s, x >) \tag{9.1}$$

be a Lagrangian, defined on $V \times V$, where V is a n-dimensional vector space over $\mathbb{R}$, $A_s \neq 0$, $s \in S$; $S \neq \emptyset$, and $< \cdot, \cdot >$ is a non-degenerate real-valued quadratic form on V. Let

$$S = S_1 \sqcup \dots \sqcup S_k, \tag{9.2}$$

all $S_i \neq \emptyset$, and

$$< b_s, b_{s'} >= 0, \tag{9.3}$$

for all $s \in S_i$, $s' \in S_j$, $i \neq j$; $i, j = 1, \ldots, k$.

Let us suppose that there exists a set $h_s \in \mathbb{R}$, $h_s \neq 0$, $s \in S$, such that

$$\sum_{s \in S} < b_s, b_{s'} > h_{s'} = -1, \tag{9.4}$$

for all $s \in S$, and

$$\frac{A_s}{h_s} = \frac{A_{s'}}{h_{s'}}, \tag{9.5}$$

$s, s' \in S_i$, $i = 1, \ldots, k$, (the ratio A_s/h_s is constant inside S_i).

Then, the Euler-Lagrange equations for the Lagrangian (9.1)

$$\ddot{x} + \sum_{s \in S} 2A_s b_s \exp(2 < b_s, x >) = 0, \tag{9.6}$$

have the following special solutions

$$x(t) = \frac{1}{2} \sum_{s \in S} h_s b_s \ln \left[y_s^2(t) \left| \frac{2A_s}{h_s} \right| \right] + \alpha t + \beta, \tag{9.7}$$

where $\alpha, \beta \in V$,

$$< \alpha, b_s >=< \beta, b_s >= 0, \tag{9.8}$$

$s \in S$, and functions $y_s(t) \neq 0$ satisfy the equations

$$\frac{d}{dt} \left(y_s^{-1} \frac{dy_s}{dt} \right) = -\xi_s y_s^{-2}, \tag{9.9}$$

with

$$\xi_s = \mathrm{sign} \left(\frac{A_s}{h_s} \right), \tag{9.10}$$

$s \in S$, and coincide inside blocks:

$$y_s(t) = y_{s'}(t), \tag{9.11}$$

$s, s' \in S_i$, $i = 1, \ldots, k$. More explicitly

$$y_s(t) = s(t - t_s, \xi_s, C_s), \tag{9.12}$$

where constants $t_s, C_s \in \mathbb{R}$ coincide inside blocks

$$t_s = t_{s'}, \qquad C_s = C_{s'}, \tag{9.13}$$

$s, s' \in S_i$, $i = 1, \ldots, k$, and

$$s(t, \xi, C) \equiv \frac{1}{\sqrt{C}} \operatorname{sh}(t\sqrt{C}), \; \xi = +1, \quad C > 0; \tag{9.14}$$

$$\frac{1}{\sqrt{-C}} \sin(t\sqrt{-C}), \; \xi = +1, \quad C < 0; \tag{9.15}$$

$$t, \; \xi = +1, \quad C = 0; \tag{9.16}$$

$$\frac{1}{\sqrt{C}} \operatorname{ch}(t\sqrt{C}), \; \xi = -1, \quad C > 0. \tag{9.17}$$

For the energy

$$E = \frac{1}{2} < \dot{x}, \dot{x} > + \sum_{s \in S} A_s \exp(2 < b_s, x >) \tag{9.18}$$

corresponding to the solution (9.7) we have

$$E = \frac{1}{2} \sum_{s \in S} C_s(-h_s) + \frac{1}{2} < \alpha, \alpha > . \tag{9.19}$$

For dual vectors $u^s \in V^*$ defined as $u^s(x) =< b_s, x >$, $\forall x \in V$, we have $< u^s, u^l >_* =< b_s, b_l >$, where $< \cdot, \cdot >_*$ is dual form on V^*. The orthogonality conditions (9.8) read

$$u^s(\alpha) = u^s(\beta) = 0, \tag{9.20}$$

$s \in S$.

References

1. V.N. Melnikov, Multidimensional Classical and Quantum Cosmology and Gravitation.Exact Solutions and Variations of Constants. CBPF-NF-051/93, Rio de Janeiro, 1993;
   V.N. Melnikov. In: *Cosmology and Gravitation*, ed. M.Novello (Editions Frontieres, Singapore, 1994) p. 147.
2. V.N. Melnikov, Multidimensional Cosmology and Gravitation, CBPF-MO-002/95, Rio de Janeiro, 1995, 210 p.
   V.N. Melnikov. In *Cosmology and Gravitation.II* ed. M. Novello (Editions Frontieres, Singapore, 1996) p. 465.
3. K.P. Staniukovich and V.N. Melnikov, *Hydrodynamics, Fields and Constants in the Theory of Gravitation*, (Energoatomizdat, Moscow, 1983), (in Russian).
4. C. Hull and P. Townsend, Unity of Superstring Dualities, *Nucl. Phys.* **B 438**, 109 (1995).
   P. Horava and E. Witten, *Nucl. Phys.* **B 460**, 506 (1996).
5. C.M. Hull, String dynamics at strong coupling, *Nucl. Phys.* **B 468**, 113 (1996).
6. J.M. Schwarz, Lectures on Superstring and M-Theory Dualities, hep-th/9607201;

7. M.J. Duff, M-theory (the Theory Formerly Known as Strings), hep-th/9608117.
8. C. Vafa, Evidence for F-Theory, hep-th/9602022; *Nucl. Phys.* **B 469**, 403 (1996).
9. H. Nicolai, On M-theory, hep-th/9801090.
10. V.N. Melnikov, *Int. J. Theor.Phys.* **33**, N7, 1569 (1994).
11. V. de Sabbata, V.N.Melnikov and P.I.Pronin, *Prog. Theor. Phys.* **88**, 623 (1992).
12. V.N. Melnikov. In: *Gravitational Measurements, Fundamental Metrology and Constants.* Eds. V. de Sabbata and V.N. Melnikov (Kluwer Academic Publ.) Dordtrecht, 1988, p.283.
13. A.J. Sanders and G.T. Gillies, *Rivista Nuovo Cim.* **19**, N2, 1 (1996).
14. A.J. Sanders and G.T. Gillies, *Grav. and Cosm.* **3**, N4(12), 285 (1997).
15. A.J. Sanders and W.E. Deeds. *Phys. Rev.* **D 46**, 480 (1992).
16. G.T. Gillies, *Rep. Progr. Phys.* **60**, 151 (1997).
17. V. Achilli et al., *Nuovo Cim.* **B 12**, 775 (1997).
18. T. Kaluza, *Sitzungsber. Preuss. Akad. Wiss. Berlin Phys. Math.*, **K1** 33, 966 (1921).
19. O. Klein, *Z. Phys.* **37**, 895 (1926).
20. V. De Sabbata and E. Schmutzer, *Unified Field Theories in more than Four Dimensions*, (World Scientific, Singapore, 1982).
21. H. C. Lee, *An Introduction to Kaluza-Klein Theories*, (World Scientific, Singapore, 1984).
22. Yu.S. Vladimirov *Physical Space-Time Dimension and Unification of Interactions* (University Press, Moscow, 1987) (in Russian).
23. P.S. Wesson and J. Ponce de Leon, *Gen. Rel. Gravit.* **26**, 555 (1994).
24. P. Jordan, Erweiterung der projektiven Relativitatstheorie, *Ann. der Phys.* 219 (1947).
25. C. Brans and R.H. Dicke, *Phys. Rev. D* **124**, 925 (1961).
26. E. Cremmer, B. Julia, and J. Scherk, *Phys. Lett.* **B76** 409 (1978).
27. A. Salam and E. Sezgin, eds., *Supergravities in Diverse Dimensions*, reprints in 2 vols., (World Scientific, Singapore, 1989).
28. M.B. Green, J.H. Schwarz and E. Witten, *Superstring Theory* (Cambridge University Press., Cambridge, 1987).
29. V.A. Belinskii and I.M. Khalatnikov, *ZhETF*, **63**, 1121 (1972).
30. P. Forgacs and Z. Horvath, *Gen. Rel. Grav.* **11**, 205 (1979).
31. A. Chodos and S. Detweyler, *Phys. Rev.* **D 21**, 2167 (1980).
32. P.G.O. Freund, *Nucl. Phys.* **B 209**, 146 (1982).
33. R. Abbot, S. Barr and S. Ellis, *Phys. Rev.* **D 30**, 720 (1984).
34. V.A. Rubakov and M.E. Shaposhnikov, *Phys. Lett.* **B 125** 136 (1983).
35. D. Sahdev, *Phys. Lett.* **B 137**, 155 (1984).
36. E. Kolb, D. Linkley and D. Seckel, *Phys. Rev.* **D 30** 1205 (1984).
37. S. Ranjbar-Daemi, A. Salam and J. Strathdee, *Phys. Lett.* **B 135**, 388 (1984).
38. D. Lorentz-Petzold, *Phys. Lett.* **B 148** 43 (1984).
39. R. Bergamini and C.A. Orzalesi, *Phys. Lett.* **B 135**, 38 (1984).
40. M. Gleiser, S. Rajpoot and J.G. Taylor, *Ann. Phys. (NY)* **160**, 299 (1985).
41. U. Bleyer and D.-E. Liebscher, in *Proc. III Sem. Quantum Gravity* ed. M.A. Markov, V.A. Berezin and V.P. Frolov (Singapore, World Scientific, 1985) p. 662.

42. U. Bleyer and D.-E. Liebscher, *Gen. Rel. Gravit.* **17**, 989 (1985).
43. M. Demianski, Z. Golda, M. Heller and M. Szydlowski, *Class. Quantum Grav.* **3**, 1190 (1986).
44. D.L. Wiltshire, *Phys. Rev.* **D 36**, 1634 (1987).
45. U. Bleyer and D.-E. Liebscher, *Annalen d. Physik (Lpz)* **44** 81 (1987).
46. G.W. Gibbons and K. Maeda, *Nucl. Phys. B* **298**, 741 (1988).
47. Y.-S. Wu and Z. Wang, *Phys. Rev. Lett.* **57** 1978 (1986).
48. G.W. Gibbons and D.L. Wiltshire, *Nucl. Phys.* **B 287**, 717 (1987).
49. V.D. Ivashchuk and V.N. Melnikov, *Nuovo Cimento* **B 102**, 131 (1988).
50. K.A. Bronnikov, V.D. Ivashchuk and V.N. Melnikov, *Nuovo Cimento* **B 102**, 209 (1988).
51. V.D. Ivashchuk and V.N. Melnikov, *Phys. Lett.* **A 135**, 465 (1989).
52. V.D. Ivashchuk, V.N. Melnikov and A.I. Zhuk, *Nuovo Cimento* **B 104**, 575 (1989).
53. V.A. Berezin, G. Domenech, M.L. Levinas, C.O. Lousto and N.D. Umerez, *Gen. Relativ. Grav.* **21**, 1177 (1989).
54. V.D. Ivashchuk and V.N. Melnikov, *Chinese Phys. Lett.* **7**, 97 (1990).
55. M. Demiansky and A. Polnarev, *Phys. Rev.* **D 41**, 3003 (1990).
56. S.B. Fadeev, V.D. Ivashchuk and V.N. Melnikov in *Gravitation and Modern Cosmology* (Plenum, N.-Y., 1991) p. 37.
57. U. Bleyer, D.-E. Liebscher and A.G. Polnarev, *Class. Quant. Grav.* **8**, 477 (1991).
58. V.D. Ivashchuk, *Phys. Lett.* **A 170**, 16 (1992).
59. A. Zhuk, *Class. Quant. Grav.* **9**, 202 (1992).
60. A. Zhuk, *Phys. Rev.* **D 45**, 1192 (1992).
61. A. Zhuk, *Sov. Journ. Nucl. Phys.* **55**, 149 (1992).
62. A.I. Zhuk, *Sov. Journ. Nucl. Phys.* **56**, 223 (1993).
63. C.W. Misner, In: Magic without Magic: John Archibald Wheeler, ed. J.R. Klauder, Freeman, San Francisko, 1972.
64. J.J. Halliwell, *Phys. Rev.* **D 38**, 2468 (1988).
65. S.W. Hawking and D.N. Page, *Phys. Rev. D* **42**, 2655 (1990).
66. H. Liu, P.S. Wesson and J. Ponce de Leon, *J. Math. Phys.* **34**, 4070 (1993).
67. V.R. Gavrilov, *Hadronic J.* **16**, 469 (1993).
68. V.D. Ivashchuk and V.N. Melnikov, *Teor. Mat. Fiz.* **98**, 312 (1994) (in Russian).
69. V.D. Ivashchuk and V.N. Melnikov, Multidimensional cosmology with m-component perfect fluid, gr-qc/ 9403063; *Int. J. Mod. Phys.* **D 3**, 795 (1994).
70. V.R. Gavrilov, V.D. Ivashchuk and V.N. Melnikov, Integrable pseudo-euclidean Toda-like systems in multidimensional cosmology with multicomponent perfect fluid, *J. Math. Phys* **36**, 5829 (1995).
71. U.Bleyer and A. Zhuk, On multidimensional cosmological models with static internal spaces, *Class. and Quantum Grav.* **12**, 89 (1995).
72. U. Bleyer and A. Zhuk, Multidimensional integrable cosmological models with negative external curvature, *Grav. and Cosmol.*, **2** 106 (1995).
73. U. Bleyer and A. Zhuk, Multidimensional integrable cosmological models with positive external space curvature, *Grav. and Cosmol.* **1**, 37 (1995).
74. U. Bleyer and A. Zhuk, Kasner-like, inflationary and steady-state solutions in multidimensional cosmology, *Astron. Nachrichten* **317**, 161 (1996).
75. A.I. Zhuk, *Sov. Journ. Nucl. Phys.* **58**, No 11 (1995).

76. J.D. Barrow and J. Stein-Schabes, *Phys. Rev.* **D 32**, 1595 (1985).
77. J. Demaret, M. Henneaux and P. Spindel, *Phys. Lett.* **B 164** 27 (1985).
78. M. Szydlowski and G. Pajdosz, *Class. Quant. Grav.* **6** (1989), 1391.
79. V.D. Ivashchuk, A.A. Kirillov and V.N. Melnikov, *Izv. Vuzov (Fizika)*, **37**, No 11 (1994) 107 (in Russian).
80. V.D. Ivashchuk, A.A. Kirillov and V.N. Melnikov, *Pis'ma ZhETF* **60** No 4 (1994), 225 (in Russian).
81. V.D. Ivashchuk and V.N. Melnikov, Billiard representation for multidimensional cosmology with multicomponent perfect fluid near the singularity, *Class. Quantum Grav.* **12**, 809 (1995).
82. A.A. Kirillov and V.N. Melnikov, Dynamics Inhomogeneouties of Metric in the Vicinity of a Singularity in Multidimensional Cosmology, *Phys. Rev.* **D 52**, 723 (1995).
A.A. Kirillov and V.N. Melnikov, On Properties of Metrics Inhomogeneouties in the Vicinity of a Singularity in K-K Cosmological Models, *Astron. Astrophys. Trans.*, **10**, 101 (1996).
83. M. Rainer, *Grav. and Cosmol.* **1**, 81 (1995).
84. M. Gasperini and G. Veneziano, *Phys. Rev.* **D 50**, 2519 (1994).
85. M. Gasperini and G. Veneziano, *Mod. Phys. Lett.* **A 8**, 701 (1993).
86. C. Angelantonj, L. Amendola, M. Litterio and F. Occhionero, String cosmology and inflation, *Phys. Rev.* **D 51**, 1607 (1995).
87. D. Kramer, *Acta Physica Polonica* **2**, F. 6, 807 (1969).
88. A.I. Legkii, in *Probl. of Grav. Theory and Elem. Particles* (Atomizdat, Moscow) **10**, 149 (1979) (in Russian).
89. D.J. Gross and M.J. Perry, *Nucl. Phys.* **B 226**, 29 (1993).
90. F.R. Tangherlini, *Nuovo Cimento* **27**, 636 (1963).
91. K.A. Bronnikov, V.D. Ivashchuk in *Abstr. VIII Soviet Grav. Conf* (Erevan, EGU, 1988) p. 156.
92. S.B. Fadeev, V.D. Ivashchuk and V.N. Melnikov, *Phys. Lett.* **A 161**, 98 (1991).
93. S.B. Fadeev, V.D. Ivashchuk and V.N. Melnikov, *Chinese Phys. Lett.* **8**, 439 (1991).
94. K.A. Bronnikov and V.N. Melnikov, *Annals of Physics (N.Y.)* **239**, 40 (1995).
95. U. Bleyer and V.D. Ivashchuk, *Phys. Lett.* **B 332**, 292 (1994).
96. V.D. Ivashchuk and V.N. Melnikov, *Class. Quantum Grav.*, **11**, 1793 (1994).
97. V.D. Ivashchuk and V.N. Melnikov, *Grav. and Cosmol.* **1**, No 3, 204 (1995).
98. V.D. Ivashchuk and V.N. Melnikov, Extremal Dilatonic Black Holes in String-like Model with Cosmological Term, *Phys. Lett.* **B 384**, 58 (1996).
99. V.A. Rubakov, *Phys. Lett.* , **B 214**, 503 (1988).
100. S. Giddings and A. Strominger, *Nucl. Phys.* **B 321**, 481 (1989).
101. A.A. Kirillov, *Pis'ma ZhETF* **55**, 541 (1992).
102. E.I. Guendelman and A.B. Kaganovich, *Phys. Lett.* **B 301**, 15 (1993).
103. T. Horigushi, *Mod. Phys. Lett.* **A 8**, 777 (1993).
104. U. Bleyer, V.D. Ivashchuk, V.N. Melnikov and A.I. Zhuk, Multidimensional classical and quantum wormholes in models with cosmological constant. gr-qc/9405020; *Nucl. Phys.* **B 429** 117 (1994).
105. V.R. Gavrilov, U. Kasper, V.N. Melnikov and M. Rainer, Toda Chains with Type A_m Lie Algebra for Multidimensional m-component Perfect Fluid Cosmology, Preprint Math-97/ Univ. Potsdam, 1997.
106. V.R. Gavrilov, V.D. Ivashchuk, and V.N. Melnikov, *Class. Quant. Grav.* **13**, 3039 (1996).

107. V.R. Gavrilov and V.N. Melnikov, *Theor. Math. Phys* **114**, N3, 454 (1998).
108. V.R. Gavrilov, V.N. Melnikov and R. Triay, Exact Solutions in Multidimensional Cosmology with Shear and Bulk Viscosity, *Class. Quant. Grav.* **14**, 2203 (1997).
V.R. Gavrilov, V.N. Melnikov and M. Novello, Exact Solutions in Multidimensional Cosmology with Bulk Viscosity, *Grav. and Cosmol.* **1**, No 2, 149 (1995).
V.R. Gavrilov, V.N. Melnikov and M. Novello, Bulk Viscosity and Entropy Production in Multidimensional Integrable Cosmology *Grav. and Cosmol.* **2**, No 4(8), 325 (1996).
109. M. Rainer and A. Zhuk, *Phys. Rev.*, **D 54** 6186 (1996).
110. V.D. Ivashchuk and V.N. Melnikov, Multidimensional Gravity with Einstein Internal spaces, hep-th/9612054; *Grav. and Cosmol.* **2**, No 3 (7), 177 (1996).
111. K.A. Bronnikov and J.C. Fabris, *Grav. and Cosmol.* **2**, No 4 (8), (1996).
112. M.J. Duff, R.R. Khuri and J.X. Lu, *Phys. Rep.* **259**, 213 (1995).
113. K.S. Stelle, Lectures on Supergravity p-Branes, hep-th/9701088. hep-th/9608117.
114. G.W. Gibbons, G.T. Horowitz and P.K. Townsend, *Class. Quant. Grav.* **12**, 297 (1995); hep-th/9410073.
115. A. Dabholkar, G. Gibbons, J.A. Harvey, and F. Ruiz Ruiz, *Nucl. Phys.* **B 340**, 33 (1990).
116. C.G. Callan, J.A. Harvey and A. Strominger, *Nucl. Phys.* **359** (1991) 611; *Nucl. Phys.* **B 367**, 60 (1991).
117. M.J. Duff and K.S. Stelle, *Phys. Lett.* **B 253**, 113 (1991).
118. G.T. Horowitz and A. Strominger, *Nucl. Phys.* **B 360**, 197 (1991).
119. R. Güven, *Phys. Lett.* **B 276**, 49 (1992); *Phys. Lett.* **B 212**, 277 (1988).
120. R. Kallosh, A. Linde, T. Ortin, A. Peet and A. van Proeyen, *Phys. Rev.* **D 46**, 5278 (1992).
121. H. Lü, C.N. Pope, E. Sezgin and K. Stelle, *Nucl. Phys.* **B 456**, 669 (1995).
122. A. Strominger, *Phys. Lett.* **B 383**, 44 (1996); hep-th/9512059.
123. P.K. Townsend, *Phys. Lett.* **B 373**, 68 (1996); hep-th/9512062.
124. A.A. Tseytlin, *Nucl. Phys.* **B 487**, 141 (1997); hep-th/9609212.
125. A.A. Tseytlin, *Mod. Phys. Lett.* **A11**, 689 (1996); hep-th/9601177.
126. G. Papadopoulos and P.K. Townsend, *Phys. Lett.* **B 380**, 273 (1996).
127. A.A. Tseytlin, Harmonic Superpositions of M-branes, hep-th/9604035; *Nucl. Phys.* **B 475**, 149 (1996).
128. J.P. Gauntlett, D.A. Kastor, and J. Traschen, Overlapping Branes in M-Theory, hep-th/9604179; *Nucl. Phys.* **B 478**, 544 (1996).
129. N. Khvengia, Z. Khvengia, H. Lü, C.N. Pope, Intersecting M-Branes and Bound States, hep-th/9605082.
130. H. Lü, C.N. Pope, and K.W. Xu, Liouville and Toda Solitons in M-Theory, hep-th/9604058.
131. M. Cvetic and A. Tseytlin, *Nucl. Phys.* **B 478**, 181 (1996).
132. I.R. Klebanov and A.A. Tseytlin, Intersecting M-branes as Four-Dimensional Black Holes, *Preprint* PUPT-1616, Imperial/TP/95-96/41, hep-th/9604166; *Nucl. Phys.* **B 475**, 164 (1996).
133. N. Ohta and T. Shimizu, Non-extreme Black Holes from Intersecting M-branes, hep-th/9701095.
134. H. Lü, C.N. Pope, and K.S.Stelle, Vertical Versus Diagonal Reduction for p-Branes, hep-th/9605082.

135. E. Bergshoeff, R. Kallosh and T. Ortin, Stationary Axion/Dilaton Solutions and Supersymmetry, hep-th/9605059; *Nucl. Phys.* **B 478**, 156 (1996).
136. G. Clément and D.V. Gal'tsov, Stationary BPS solutions to dilaton-axion gravity *Preprint* GCR-96/07/02 DTP-MSU/96-11, hep-th/9607043.
137. A. Volovich, Three-block p-branes in various dimensions, hep-th/9608095.
138. I.Ya. Aref'eva and A.I. Volovich, Composite p-branes in Diverse Dimensions, *Preprint* SMI-19-96, hep-th/9611026; *Class. Quantum Grav.* 14 (11), 2990 (1997).
139. I.Ya. Aref'eva, K. Viswanathan, A.I. Volovich and I.V. Volovich, p-Brane Solutions in Diverse Dimensions, hep-th/9701092.
140. N. Khvengia, Z. Khvengia, H. Lănd C.N. Pope, Toward Field Theory of F-Theory, hep-th/9703012.
141. V.D. Ivashchuk and V.N. Melnikov, Intersecting p-Brane Solutions in Multidimensional Gravity and M-Theory, hep-th/9612089; *Grav. and Cosmol.* **2**, No 4, 297 (1996).
142. V.D. Ivashchuk and V.N. Melnikov, *Phys. Lett. B* **403**, 23 (1997).
143. V.D. Ivashchuk and V.N. Melnikov, Sigma-Model for Generalized Composite p-branes, hep-th/9705036; *Class. and Quant. Grav.* **14**, 11, 3001 (1997).
144. V.D. Ivashchuk, M. Rainer and V.N. Melnikov, Multidimensional Sigma-Models with Composite Electric p-branes, gr-qc/9705005; *Gravit. and Cosm.* **4**, No1 (13) (1998).
145. E. Bergshoeff, M. de Roo, E. Eyras, B. Janssen and J.P. van der Schaar, hep-th/9612095.
146. I.Ya. Aref'eva and O.A. Rytchkov, Incidence Matrix Description of Intersecting p-brane Solutions, hep-th/9612236.
147. R. Argurio, F. Englert and L. Hourant, Intersection Rules for p-branes, hep-th/9701042.
148. I.Ya. Aref'eva M.G. Ivanov and O.A. Rytchkov, Properties of Intersecting p-branes in Various Dimensions, hep-th/9702077.
149. I.Ya. Aref'eva, M.G. Ivanov and I.V. Volovich, Non-Extremal Intersecting p-Branes in Various Dimensions, hep-th/9702079; *Phys. Lett.* **B 406**, 44 (1997).
150. N. Ohta, Intersection Rules for Non-extreme p-branes, hep-th/9702164.
151. K.A. Bronnikov, V.D. Ivashchuk and V.N. Melnikov, The Reissner-Nordström Problem for Intersecting Electric and Magnetic p-Branes, gr-qc/9710054; *Grav. and Cosmol.* **3**, No 3 (11), 203 (1997).
152. K.A. Bronnikov, U. Kasper and M. Rainer, Intersecting Electric and Magnetic p-Branes: Spherically Symmetric Solutions, gr-qc/9708058.
153. K.A. Bronnikov, M.A. Grebeniuk, V.D. Ivashchuk and V.N. Melnikov, Integrable Multidimensional Cosmology for Intersecting p-branes, *Grav. and Cosmol.* **3**, No 2(10), 105 (1997).
154. M.A. Grebeniuk, V.D. Ivashchuk and V.N. Melnikov, Integrable Multidimensional Quantum Cosmology for Intersecting p-Branes, *Grav. and Cosmol.* **3**, No 3 (11), 243 (1997), gr-qc/9708031.
155. H. Lü, J. Maharana, S. Mukherji and C.N. Pope, Cosmological Solutions, p-branes and the Wheeler De Witt Equation, hep-th/9707182.
156. H. Lü, S. Mukherji, C.N. Pope and K.-W. Xu, Cosmological Solutions in String Theories, hep-th/9610107.
157. S. Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989).
158. V.D. Ivashchuk and V.N. Melnikov, Multidimensional Ouantum Cosmology with Intersecting p-branes, *Hadronic J.* **21**, 319 (1998).

159. M.A. Grebeniuk, V.D. Ivashchuk and V.N. Melnikov, Multidimensional Cosmology for Intersecting p-branes with Static Internal Spaces, *Grav. and Cosm.*, **4**, No 2(14) (1998).
160. S.D. Majumdar, *Phys. Rev.* **72**, 930 (1947);
A. Papapetrou, *Proc. R. Irish Acad.* **A51**, 191 (1947).
161. N.M. Bocharova, K.A. Bronnikov and V.N. Melnikov, *Vestnik MGU (Moscow Univ.)*, **6**, 706 (1970)(in Russian) - first MP-type solution with conformal scalar field;
K.A. Bronnikov, *Acta Phys. Polonica* , **B4**, 251 (1973);
K.A. Bronnikov and V.N. Melnikov, in *Problems of Theory of Gravitation and Elementary Particles* , **5**, 80 (1974) (in Russian) - first MP-type solution with conformal scalar and electromagnetic fields.
162. M. Szydłowski, *Acta Cosmologica* **18**, 85 (1992).
163. G.W. Gibbons and S.W. Hawking, *Phys. Rev.* **D 15**, 2752 (1977).
164. V.D. Ivashchuk and V.N. Melnikov, *Int. J. Mod. Phys. D* **4**, 167 (1995).
165. V.D. Ivashchuk and V.N. Melnikov, On Singular Solutions in Multidimensional Gravity, hep-th/9612089; *Grav. and Cosmol.* **1**, No 3, 204 (1996).
166. V.D. Ivashchuk and V.N. Melnikov, Multidimensional Classical and Quantum Cosmology with Intersecting p-branes, hep-th/9708157; *J. Math. Phys.*, **39**, 2866 (1998).
167. K.A. Bronnikov, Block-orthogonal Brane systems, Black Holes and Wormholes, hep-th/9710207; *Grav. and Cosmol.* **4**, No 2 (14), (1998).
168. V.D. Ivashchuk and V.N. Melnikov, Mudjumdar-Papapetrou Type Solutions in Sigma-model and Intersecting p-branes, hep-th/9702121; *Class. Quantum Grav.***16**, 849 (1999)
169. V.D.Ivashchuk, Composite p-branes on Product of Einstein spaces, *Phys. Lett.* **B 434**, 28 (1998).
170. T. Damour, “Gravitation, Experiment and Cosmology”, gr-qc/9606079.
171. R.D. Reasenberg et. al., *Astrophys. J.* **234**, L219 (1979).
172. J.O. Dickey et al., *Science* **265**, 482 (1994).
173. K. Nordtvedt, *Phys. Rev.* **169**, 1017 (1968).

Open Issues

- A Anderson:
  1. **Question:** Does there exist a variational formulation of hyperbolic formulations of general relativity?
  2. **Question:** Is the metric the right variable to quantise in quantum gravity? If not, what is?
  3. **Question:** Do there exist nontrivial strong field effects in black hole collisions?
  4. **Question:** Develop a perturbation theory for the intermediate stages of binary collapse and connect the post-Newtonian regime with the strong-field regime.
     Construct realistic astrophysical data for numerical evolution.
     Comment: J. York: Investigate sensitivity to initial conditions.
- J Barrow:
  1. **Question:** Characterise parts of general classical cosmological solutions (in vacuum or with a perfect fluid) as $t \to \pm\infty$; consider the constraints imposed by a nontrivial topology. In particular, does there exist a nondecreasing functional like the area of event horizon? Is there an infinite number of curvature oscillations?
  2. **Question:** Define a black hole in a nonasymptotically flat universe.
     Comment: T Jacobson: If there is no future singularity, take the boundary of $I^-(\gamma)$.
  3. **Question:** Prove that closed universes recollapse (with spherical symmetry??) in vacuum or with a perfect fluid satisfying the strong energy condition. Alternatively, provide necessary and sufficient conditions under which recollapse always occurs.
  4. **Question:** Find necessary and sufficient conditions for the computational equivalence problem for metrics to be Gödel undecidable.
- R Beig:
  1. **Question:** Prove uniqueness of rotating black holes without unjustified analyticity assumptions.
- J G Cardoso:
  1. **Question:** Find out whether one might develop a theory of wave mappings in spacetimes with torsion. If so, give a physical interpretation.
- Y Choquet-Bruhat:

S. Cotsakis and G.W. Gibbons (Eds.): Proceedings 1998, LNP 537, pp. 249–251, 2000.

1. **Question:** Pass from Sobolev spaces to those admitting less regularity. Prove existence and uniqueness theorems. (Include models with lumps of matter.)
   Comment: S W Hawking: Isn't Nature C^{∞} or analytic? (Euclidean quantum gravity works by analytic continuation.)
   Reply: Y Choquet-Bruhat: It depends on the scale; it may be true at micro scales, but at large scales it seems that discontinuities occur.
2. **Question:** How does one pass between scales?
3. **Question:** Construct solutions of constraints with large variation of $\mathrm{Tr}K_{ij}$.
   Comment: It is easy to find them out with matter (e.g., expanding and contracting universes). Can this be done in *vacuo*?

- T Christodoulakis:
  1. **Question:** Does there exist a mathematical relation between classical time reparameterisation and the hyperbolic nature of the Wheeler-DeWitt equation? (Also with regard to the lack of L^2 property.)
     Comment: A Anderson: There exist classical solutions that do not respect the Wheeler-DeWitt cone.
- G W Gibbons:
  1. Examine the Dirichlet and Neumann problems for Euclidean Einstein equations:
     Question: Given a three-dimensional surface Σ_3 and either a metric g_{ij} (first fundamental form) or the extrinsic curvature K_{ij} (second fundamental form) on this surface, does there exist a four-dimensional manifold M_4 such that its boundary be precisely Σ_3? If so, is it unique?
     Problem: In the case of the Dirichlet problem: how does the convexity of the boundary affect the uniqueness?
  2. **Question:** Extend and generalise uniqueness theorems for black holes to higher dimensions and for p-branes.
- S W Hawking:
  1. **Question:** What is the nature of gravitational entropy? Does it lead to information loss?
  2. **Question:** Does the Bekenstein boundary lead to UV cutoff making the theory finite?
  3. **Question:** Find a realistic M-theory cosmology.
  4. **Question:** Understand the mechanism of SUSY breaking.
  5. **Question:** What happens to somebody who falls into a large black hole?
  6. **Question:** How does one describe and calculate the final disappearance of an evaporating black hole?
  7. **Question:** How to distinguish TIPs inside the big crunch from those which escape?
     Comment: T Jacobson: Consider the case without big crunch.

- T Ilmanen:
  1. **Question:** Clarify the geometry and variational properties of mean convex and marginally trapped closed 2-surfaces.
  2. **Question:** Interrelate the mass proofs.
  3. **Question:** Minimise ADM mass by extending a fragment Ω of the manifold.
  4. **Question:** Construct vacuum examples of cosmic censorship violation; is it easier in higher dimensions?
- T Jacobson:
  1. **Question:** Does there exist a Regge calculus (classical) resolution of Schwarzschild singularities?
     Comment: G W Gibbons: Find the Regge calculus analogue of ADM mass.
  2. **Question:** Is the generalised second law true with higher derivatives?
  3. **Question:** How general are the Penrose inequalities?
     Comment: G W Gibbons: State and prove as many as you can.
- Z Perjes:
  1. **Question:** Develop the current post-Newtonian formalisms up to a point where one could treat binary star coalescence.
  2. **Question:** Get a rotating star solution in the exact solutions book.
- I Racz:
  1. **Question:** Give a clean mathematical formulation of the possible final states of gravitational collapse and states which are attained (cf. Christodoulou in the spherically symmetric collapse with a scalar field.)
- G Savvidis:
  1. **Question:** Given that geometry and metric fluctuate, why does Rip van Winkle[1] awake to find his friends still discussing quantum gravity?
- S Cotsakis:
  1. **Question:** Does there exist a description of black hole entropy in terms of membranes near the horizon, for *all* types of black holes?
- J York:
  1. **Question:** On behalf of K Kuchar: What is the nature of time in quantum cosmology?
     Addendum: J York: Is this the correct question?
  2. **Question:** Will a satisfactory theory of quantum gravity allow us to retain a useful concept of mass-energy?
  3. **Question:** Will formulating quantum gravity teach us something new about quantum mechanics?
  4. **Question:** Can a black hole be described entirely in terms of quasi-normal modes?

[1] Recall the story of Rip van Winkle who slept from the Dutch period into the days of George Washington, a period of some two hundred years.

List of Contributors

Arlen Anderson
Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC 27599-3255, USA

R. Beig
Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A–1090 Wien, Austria

Dieter Brill
University of Maryland, College Park, MD 20742, USA and
Albert-Einstein-Institut, Potsdam, Germany

Yvonne Choquet-Bruhat
Gravitation et Cosmologie Relativiste, t.22-12, Université Paris VI,
75252 Paris, France

Arthur E. Fischer
Department of Mathematics, University of California,
Santa Cruz, CA 95064, USA

G. W. Gibbons
D.A.M.T.P., Cambridge University, Silver Street,
Cambridge, CB3 9EW, U.K.

S.W. Hawking
D.A.M.T.P., Cambridge University, Silver Street,
Cambridge, CB3 9EW, U.K.

V.D. Ivashchuk
Center for Gravitation and Fundamental Metrology VNIIMS,
3-1 M. Ulyanovoy Str.,
Moscow, 117313, Russia

V.N. Melnikov
Center for Gravitation and Fundamental Metrology VNIIMS,
3-1 M. Ulyanovoy Str.,
Moscow, 117313, Russia

Vincent Moncrief
Departments of Mathematics and Physics, Yale University,
New Haven, CT 06511, USA

Hernando Quevedo
Instituto de Ciencias Nucleares, UNAM, A. Postal 70-543,
México 04510 D.F., Mexico

Michael P. Ryan, Jr.
Instituto de Ciencias Nucleares, UNAM, A. Postal 70-543,
México 04510 D.F., Mexico

James W. York Jr.
Department of Physics and Astronomy, University of North Carolina,
Chapel Hill, NC 27599-3255, USA