Commun. Math. Phys. 151, 53–66 (1993)

Communications in
Mathematical
Physics

On the Topology of Black Holes*

Gregory J. Galloway
Department of Mathematics and Computer Science, University of Miami, Coral Gables, FL 33124, USA

Received March 19, 1992

Dedicated to Ted Frankel

Abstract. We establish from local hypotheses some results concerning the final state topology of black holes. We show that the surface of a black hole must have 2-sphere topology and that the topology of space in its vicinity is simple.

1. Introduction

The classical theory of black holes (as elucidated, for example, in [HE]) imposes at the outset numerous global asymptotic conditions. For instance, spacetime is required to be weakly asymptotically simple and empty (which implies the existence of past and future null infinity $\mathscr{I}^{\pm}$) and future asymptotically predictable from a partial Cauchy surface S (which asserts that cosmic censorship holds). Certain global topological assumptions are imposed, as well (cf. [HE], p. 317). Although the purpose of these conditions is to model the region of spacetime in the vicinity of an isolated, or quasi-isolated, gravitating system, mathematically, they are conditions on all of spacetime. Since the formation of a black hole due to the gravitational collapse of some stellar object is viewed as a rather local phenomenon, it would seem to be of interest to consider what properties of black holes can be derived from purely local hypotheses (i.e., from assumptions made on an arbitrarily small region of spacetime in the vicinity of the black hole). The aim of this paper is to establish from local hypotheses some results concerning the final state topology of black holes. We show that, in the steady state limit, the surface of a black hole must have 2-sphere topology (thus recovering Hawking's result ([HE], p. 325) in the static case from local hypotheses) and that the region of space in the vicinity of the black hole is correspondingly simple. Before giving a precise statement of these results we briefly describe the philosophy behind them. Numerous examples and theorems (e.g. [G, L, FG, MA]) indicate that nontrivial spatial topology cannot support a state of gravitational equilibrium, and, in fact, tends to

* This work has been partially supported by NSF grants DMS-8802877 and DMS-9006678

induce gravitational collapse. In the standard collapse scenario, the collapsing region crosses an event horizon which forms during the creation of a black hole, and the region outside the event horizon settles down into a steady state. One then expects the topology outside the event horizon to be simple, for, otherwise, further gravitational collapse would ensue. This reasoning suggests a notion of "topological censorship," in which nontrivial topology becomes hidden behind the event horizon. We proceed to a detailed description of our results.

Let V be a compact 3-dimensional Riemannian manifold with boundary ∂V isometrically imbedded as a spacelike hypersurface into a spacetime M. Assume $\partial V = \partial^H \cup \partial^E$, where ∂^H (the "black hole boundary") consists of $k \geqq 0$ compact, possibly nonorientable surfaces and ∂^E (the "external boundary" surrounding the black holes) is orientable with at least one component topologically a 2-sphere. We make some comments about these assumptions.

(1) Physically, V is to be thought of as a finite region of space (at some instant of time) surrounding (but exterior to) k black holes. Since we will permit the presence of arbitrary matter and electrostatic fields and do not make any asymptotic assumptions, we allow for the possibility of multiple black holes in equilibrium.

(2) In the most direct interpretation of our model, we think of ∂^E as consisting of a single 2-sphere surrounding the k black holes. However there are a couple of reasons for allowing a priori ∂^E to have more than one component. One reason is that the method of proof used to rule out nonorientable black holes forces us to consider the situation in which ∂^E is not connected. Another reason is that we can, if we like, think of the components of ∂^E as corresponding to the ends of space. Our results then have the added benefit of ruling out multiple ends.

In order to study the final state topology of the exterior region V, we take the usual approach of invoking the existence of a timelike Killing vector field X defined, in our situation, in a neighborhood of $V \setminus \partial^H$ in M. We make the additional simplifying assumption that X is orthogonal to $V \setminus \partial^H$. By conservation of vorticity, it follows that X is actually irrotational (i.e., hypersurface orthogonal) in a neighborhood of $V \setminus \partial^H$, and hence that M is *static* in this neighborhood. A study of the more general *stationary* case (in which X has nonzero rotation) is deferred to the future.

By introducing coordinates adapted to the vector field X, the metric of M near $V \setminus \partial^H$ can be written in standard static form as,

$$ds^2 = -\phi^2 dt^2 + d\sigma^2 , \tag{1}$$

where $X = \frac{\partial}{\partial t}$, $\phi = \sqrt{-\langle X, X \rangle}$ is independent of t, and $d\sigma^2$ is the induced Riemannian metric on V. We assume that ϕ satisfies the following standard (for static black hole spacetimes) regularity conditions.

(1) $\phi : V \setminus \partial^H \to \mathbb{R}$ extends smoothly to ∂^H, and
(2) $\phi = 0$ and $d\phi \neq 0$ along ∂^H.

In the static case, the notions of event horizon, apparent horizon, and stationary limit surface (or infinite red shift surface) coincide. Hence, in the static case, the vanishing of ϕ becomes the defining condition for the black hole boundary. The static field equations (specifically Eq. (9) in Sect. 3) and the vanishing of ϕ along ∂^H imply that ∂^H is totally geodesic, a fact that will be used in Sect. 2.

A compact Riemannian manifold with boundary V which satisfies the conditions of the previous paragraphs shall be referred to as a *static body surrounding k blackholes*. In our main theorem it is assumed that the external boundary ∂^E of the body V is mean convex, i.e., has negative mean curvature with respect to the outward pointing normal, $H_{\partial^E} < 0$. (We use the sign convention in which the mean curvature is *minus* the divergence of the outward pointing normal.) This mean convexity assumption can be thought of as a mild (and realistic) asymptotic flatness condition. Let us say that V is *standard* if it has mean convex external boundary. The truncated Flamm paraboloid consisting of all points in the slice $t = 0$ in Schwarzschild spacetime with $2m \leqq r \leqq r_0$ $(r_0 > 2m)$ is an example of a standard static body surrounding one black hole. A similar example occurs in the Reissner–Nordstrom solution.

The following theorem shows that, provided appropriate energy conditions are satisfied, the topology of a standard static body surrounding k black holes is as simple as possible.

Main Theorem. *Let V be a standard static body in M surrounding $k \geqq 0$ black holes; hence, in particular, $\partial V = \partial^H \cup \partial^E$, where ∂^H has k components, and ∂^E is mean convex, orientable and has at least one 2-sphere component. Assume that the energy condition* (EC1), *or, more generally, the energy condition* (EC2) (*stated below*) *is satisfied. Then ∂^E is connected* (*and hence consists of a single 2-sphere*), *each component of ∂^H is a 2-sphere, and V is diffeomorphic to a closed 3-ball minus k open 3-balls.*

EC1 (*matter fields only*). For each nonzero null vector K along V,

$$\mathrm{Ric}(K, K) \geqq 0 \text{ and } = 0 \text{ iff } \mathrm{Ric} = 0 \text{ (i.e., the full Ricci tensor vanishes)}\,,$$

where $\mathrm{Ric} = \mathrm{Ric}_M$ is the Ricci tensor of spacetime.

In the case of a perfect fluid with mass density ρ and isotropic pressure p, EC1 becomes: $\rho + p \geqq 0$ and $= 0$ iff $\rho = p = 0$. EC1 is a perfectly reasonable energy condition for ordinary matter fields having nonnegative stresses. However for fields with negative stresses, such as an electrostatic field, EC1 does not in general hold. (For such fields, $\mathrm{Ric}(K, K) = 0$ does not necessarily imply $\mathrm{Ric} = 0$). The energy condition EC2 applies to the case in which both matter fields and an electrostatic field may be present. In order to state EC2, it is more convenient to make explicit use of the Einstein equation.

EC2 (*matter fields plus electrostatic field*). The Einstein equation,

$$\mathrm{Ric} - \frac{1}{2}Rg = 8\pi\mathscr{T} \tag{2}$$

holds in a neighborhood of $V\backslash\partial V$ in M. The energy-momentum tensor $\mathscr{T}$ consists of two parts, $\mathscr{T} = \mathscr{M} + \mathscr{E}$, where $\mathscr{M}$, the energy-momentum tensor for the matter fields and $\mathscr{E}$, the energy-momentum tensor for the electrostatic field, satisfy the following conditions.

(1) $\mathscr{M}$ obeys EC1, i.e., for each nonzero null vector K along V, $\mathscr{M}(K, K) \geqq 0$ and $= 0$ iff $\mathscr{M} = 0$.

(2) $\mathscr{E}$ is related to the electromagnetic field tensor $\mathscr{F}$ in the usual way,

$$E_{ij} = \frac{1}{4\pi}\left(F_i^{\ k} F_{jk} - \frac{1}{4} F_{rs} F^{rs} g_{ij} \right), \tag{3}$$

and $\mathscr{F}$ is such that observers comoving with the Killing field X measure a time independent electric field $\vec{E}$ and zero magnetic field, $\vec{B} = 0$.

(3) If $\mathscr{M}$ vanishes in some region of V then $\mathscr{F}$ obeys the *free space* Maxwell equations in that region.

Remark. In physical terms, condition (3) of EC2 simply requires that particles in V carrying charge also carry gravitational mass.

The proof of the theorem consists of a topological lemma and a variational argument. The topological lemma, which is presented in Sect. 2, shows that if a standard static body V surrounding k black holes does not have the desired simple topology then there exists in $V \backslash \partial V$ a minimal surface Σ which has least area in a certain class. The proof of this lemma relies heavily on various aspects of the paper of Meeks, Simon and Yau [MSY] concerning minimizing area in isotopy class. Section 2 includes a summary of the relevant parts of their work. In Sect. 3 a variational argument is used to show that if EC1 holds then $V \backslash \partial V$ cannot contain such least area surfaces Σ. In Sect. 4 this argument is extended to the case in which EC2 holds. For reasons discussed in Sect. 4, the proof in the electrostatic case is rather more complicated.

The author is indebted to Leon Simon for several helpful conversations during the course of this work. The author would also like to express his thanks to Ted Frankel for helpful discussions concerning the extension to the electrostatic case. Part of the work on this paper was carried out during a visit to the Centre for Mathematical Analysis in Canberra. The author wishes to express his thanks to the Centre for its hospitality and financial support.

2. The Topological Lemma

Our work relies on several aspects of the paper of Meeks, Simon and Yau [MSY] concerning the problem of minimizing area in isotopy class, namely, the fundamental result establishing the existence and regularity of a minimizer for a given isotopy class, the description of the topological relationship between the minimizer and the elements of the isotopy class, and the characterization of handlebodies. We begin this section by paraphrasing these results, tailoring them, when convenient, to our needs. Later in the section we will use these results to prove what we refer to as the topological lemma.

Let N be a compact 3-dimensional Riemannian manifold with boundary ∂N (possibly empty) which is mean convex, $H_{\partial N} < 0$. Let S be a connected two-sided compact surface in $N \backslash \partial N$, and let $\mathscr{I}(S)$ denote the (ambient) isotopy class of S. (A surface is two-sided if it admits a smooth unit normal.) Suppose,

$$\inf_{S' \in \mathscr{I}(S)} A(S') = \delta > 0 ,$$

where A = area. Let $\{\Sigma_k\}$ be a minimizing sequence for $\mathscr{I}(S)$, i.e. $\{\Sigma_k\} \subset \mathscr{I}(S)$ and $\lim_{k \to \infty} A(\Sigma_k) = \delta$. Then MSY establish the existence of a subsequence $\{\Sigma_{k'}\}$ which converges in a suitable sense to a finite collection of compact minimal surfaces

$\Sigma^{(1)}, \ldots, \Sigma^{(R)}$ (none of which need be in $\mathscr{I}(S)$) with positive multiplicities $n_1, \ldots, n_R$, such that each $\Sigma^{(j)}$ has least area in an appropriate class. For example, if $\Sigma^{(j)}$ is two-sided then its area is less than or equal to the area of any surface isotopic to it provided the isotopy leaves $\Sigma^{(k)}$, $k \neq j$, fixed. If $\Sigma^{(j)}$ is one-sided then it also minimizes area, but in a slightly more complicated sense (cf. [MSY], p. 635, p. 645). Since the one-sided case can occur, let us state explicitly the area minimizing property of $\Sigma^{(j)}$ we shall require.

Let Σ be a compact surface in N. Recall, if Σ is one-sided, there is a standard two-sheeted covering $(\tilde{N}, \pi)$ of N such that $\tilde{\Sigma} = \pi^{-1}(\Sigma)$ is a double covering of Σ and $\tilde{\Sigma}$ is two-sided in $\tilde{N}$. Let us say that Σ is *locally of least area* provided the following holds.

(1) If Σ is two-sided then for any normal variation $u \to \Sigma_u$ of Σ, Σ satisfies: $A(\Sigma) \leqq A(\Sigma_u)$ for all u sufficiently small.

(2) If Σ is one-sided then for any normal variation $u \to \tilde{\Sigma}_u$ of the two-sided double cover $\tilde{\Sigma} \subset \tilde{N}$, with each $\tilde{\Sigma}_u$ to one side of $\tilde{\Sigma}$, $\tilde{\Sigma}$ satisfies: $A(\tilde{\Sigma}) \leqq A(\tilde{\Sigma}_u)$ for all u sufficiently small.

Each $\Sigma^{(j)}$ is locally of least area in this sense.

Due to "bubbling off", the minimizer (as realized by the $\Sigma^{(j)}$'s) will not, in general, belong to the isotopy class $\mathscr{I}(S)$. However, by their minimization procedure, MSY are able to provide a detailed description of the topological relationship between the minimizer and S. For example, they obtain a relationship between the genera of the $\Sigma^{(j)}$'s and the genus of S, which when the $\Sigma^{(j)}$'s are all two-sided becomes,

$$\sum_{j=1}^{R} n_j \operatorname{genus}(\Sigma^{(j)}) \leqq \operatorname{genus}(S)\,. \tag{4}$$

MSY ([MSY], Proposition 1, p. 650) obtain a useful geometric characterization of handlebodies. A handlebody is a 3-dimensional manifold with boundary which is diffeomorphic to a handlebody in $\mathbb{R}^3$, i.e., a solid in $\mathbb{R}^3$ bounded by a compact surface of arbitrary genus.

Lemma 1 ([MSY]). *A 3-dimensional compact Riemannian manifold with boundary is a handlebody if and only if the isotopy class of some surface S parallel to a boundary component contains surfaces of arbitrarily small area. (Take "parallel" to mean nudged in slightly along the normal geodesics to the boundary component).*

We are now prepared to state and prove the topological lemma.

Lemma 2 (The topological lemma). *Let N be a 3-dimensional orientable compact Riemannian manifold with boundary $\partial N = \partial^A \cup \partial^B$ such that*

(1) *∂^B is totally geodesic and has $k \geqq 0$ components.*

(2) *∂^A is mean convex, $H_{\partial^A} < 0$, and has at least one component diffeomorphic to S^2.*

Then either there is a compact minimal 2-sphere (or projective plane) Σ contained in $N \backslash \partial N$ which is locally of least area as described above, or else ∂^A and each component of ∂^B are 2-spheres, and N is diffeomorphic to a closed 3-ball minus k open 3-balls.

Comment. The MSY minimizing procedure applied to a surface parallel to a component of ∂^A can produce a minimizer contained in ∂^B. The example of the truncated Flamm paraboloid mentioned in Sect. 1 clearly illustrates this possibility. The essential point of Lemma 2 for our purposes is that unless the topology of N is sufficiently simple, the MSY minimizing procedure will produce a minimizer which is contained in the *interior* of N.

Proof. If $\partial^B = \emptyset$ then Lemma 2 follows directly from Lemma 1 above and Theorem 1′ (the main existence result for manifolds with mean convex boundary) in [MSY]. Henceforth assume $\partial^B \neq \emptyset$. Let ∂_i^A, $i = 0, \ldots, l$ (with $\partial_0^A \approx S^2$) and ∂_j^B, $j = 1, \ldots, k$ denote the components of ∂^A and ∂^B, respectively. By adding a collar to N along ∂^B we can isometrically imbed N into a compact Riemannian manifold $\hat{N}$ with boundary $\partial \hat{N} = \partial^A \cup \partial^C$ with the following properties.

(1) $\partial \hat{N}$ is (strictly) mean convex, $H_{\partial \hat{N}} < 0$.

(2) $\hat{N} \setminus N$ is diffeomorphic, via the normal exponential map along ∂^B, to $[0, \varepsilon] \times \partial^B$.

(3) Any compact minimal surface in $\hat{N}$ is contained in N.

Indeed, one can take the metric on $\hat{N} \setminus N \approx [0, \varepsilon] \times \partial^B$ to be the warped product $-dt^2 \oplus G(t)^2 h$, where h is the induced metric on ∂^B and G is the function: $G(0) = 1$ and $G(t) = 1 + e^{-1/t}$, $t > 0$. The fact that ∂^B is totally geodesic guarantees that the metric $\hat{g}$ of $\hat{N}$ is $C^{1,1}$ in a neighborhood of ∂^B and C^∞ on $\hat{N} \setminus \partial^B$. This degree of regularity is sufficient for the arguments of MSY to hold (cf. [M-A]). Moreover, for each $t \in (0, \varepsilon]$, the level surface $\{t\} \times \partial^B$ is convex, and hence the maximum principle for hypersurfaces (cf. [E] for a nice exposition) ensures that property (3) holds.

Push the boundary component ∂_0^A inward slightly to obtain a 2-sphere $\partial \subset \hat{N} \setminus \partial \hat{N}$, and consider the isotopy class $\mathscr{I}(\partial, \hat{N})$ of ∂ in $\hat{N}$. Since $\hat{N}$ is not a handlebody, Lemma 1 implies,

$$\inf_{S \in \mathscr{I}(\partial, \hat{N})} A(S) = \delta > 0 . \tag{5}$$

We now carry out the MSY minimizing procedure on ∂. (Strictly speaking, this minimizing procedure takes place in a certain homogeneously regular extension of $\hat{N}$ (cf. the proof of Theorem 1′ in [MSY]). However, by the properties of this extension, there is no loss of generality in assuming the procedure takes place in $\hat{N}$.) Then, as described above, there exists a minimizing sequence $\{\Sigma_k\}$ for $\mathscr{I}(\partial, \hat{N})$ which converges to a collection of minimal surfaces $\Sigma^{(j)}$, $j = 1, \ldots, R$, each of which is locally of least area. Moreover, by the inequality (1.4) in [MSY] (which generalizes (4) above to include the one-sided case), each $\Sigma^{(j)}$ is either a 2-sphere or a projective plane. By property (3) of $\hat{N}$, each $\Sigma^{(j)}$ is contained in $N \setminus \partial^A$. If any $\Sigma^{(j)}$ is contained in $N \setminus \partial N$ we are done.

So suppose each $\Sigma^{(j)}$ meets ∂^B. Then, by the maximum principle for hypersurfaces, $\Sigma^{(j)} \subset \partial^B$ for $j = 1, \ldots, R$. Thus, relabeling if necessary, we have,

$$\Sigma^{(j)} = \partial_j^B, \quad j = 1, \ldots, R .$$

In particular, each $\Sigma^{(j)}$ is two-sided, and hence, by inequality (4), a 2-sphere.

We now make use of the more detailed information in [MSY] concerning the relationship between ∂ and the $\Sigma^{(j)}$'s. It follows from the discussion on p. 365 in [MSY] that there exists a surface $S = \bigcup_{j=1}^R S^{(j)}$ in $\hat{N}$ which satisfies the following.

(1) S is a disjoint union of spheres. More explicitly, $S^{(j)} = \bigcup_{r=1}^{n_j} S_r^{(j)}$ (disjoint union), where each $S_r^{(j)}$ is parallel to $\Sigma^{(j)}$. In fact each $S_r^{(j)}$ is obtained from $\Sigma^{(j)}$ by pushing $\Sigma^{(j)}$ along its normal geodesics an arbitrarily small amount; in particular, each $S_r^{(j)}$ is a 2-sphere.

(2) By cutting out arbitrarily small disks in S and attaching arbitrarily thin tubes, one can obtain a surface $S' \in \mathscr{I}(\partial, \hat{N})$.

We proceed to show that $\hat{N}$ has the desired simple topology. Each component of $\partial\hat{N} = \partial^A \cup \partial^C$ is an orientable surface of some genus g. Smoothly attach a handlebody with the appropriate number of handles to each component of $\partial\hat{N}$ except ∂_0^A to obtain a Riemannian manifold $\hat{N}'$ with boundary ∂_0^A. We show that $\mathscr{I}(\partial, \hat{N}')$ contains surfaces of arbitrarily small area. Note that, for each $j = 1, \ldots, R$, $\partial_j^C \approx \partial_j^B \approx S^2$ bounds a 3-cell (i.e., a manifold diffeomorphic to a closed 3-ball) in $\hat{N}'$. Since for each $r = 1, \ldots, n_j$, $S_r^{(j)}$ is parallel to ∂_j^C, $S_r^{(j)}$ bounds a 3-cell in $\hat{N}'$, as well. In fact, there exists a 3-cell B_j bounded by the "outermost" $S_r^{(j)}$ (i.e., the $S_r^{(j)}$ furthest from ∂_j^C). By shrinking B_j down "radially," $S^{(j)} = \bigcup_{r=1}^{n_j} S_r^{(j)}$ can be shrunk down isotopically to a surface $\tilde{S}^{(j)}$ of arbitrarily small area. Thus there exists a surface $\tilde{S} = \bigcup_{j=1}^{R} \tilde{S}^{(j)} \in \mathscr{I}(S, \hat{N}')$ of arbitrarily small area. From property (2) of S, one can attach arbitrarily thin tubes to $\tilde{S}$ to obtain a surface in $\mathscr{I}(\partial, \hat{N}')$ of arbitrarily small area. Hence, by Lemma 1, $\hat{N}'$ is a handlebody. In fact, since $\partial\hat{N}' \approx S^2$, $\hat{N}'$ is a 3-cell.

Thus, $\hat{N}$ is a 3-cell minus the interiors of $l + k$ handlebodies, where $l \geqq 0$ and $k \geqq R$. Again, by property (2) of S, we can attach arbitrarily thin tubes to S to obtain a sphere $\bar{S}$ isotopic to ∂. ("Arbitrarily thin" means, in particular, that the tube together with the disks that get cut out form a sphere which bounds a ball.) Each of the components of S surrounds one of the boundary spheres ∂_j^C, $j = 1, \ldots, R$, in $\hat{N}$, and does not surround any other component of $\partial\hat{N}$. Hence, $\bar{S}$ can surround any of the boundary components $\partial_1^C, \ldots, \partial_R^C$, but no others. ∂, on the other hand, surrounds the components $\partial_1^A, \ldots, \partial_l^A, \partial_1^C, \ldots, \partial_k^C$ (i.e., all of the components other than ∂_0^A). It is clear from elementary homological considerations that, under these circumstances, the only way $\bar{S}$ can be isotopic to ∂ is if $k = R$, i.e., $\partial^C = \bigcup_{j=1}^{R} \partial_j^C$ and $l = 0$, i.e., $\partial^A = \partial_0^A$. Thus, $\hat{N}$ is a closed 3-ball minus k open 3-balls. Since N is diffeomorphic to $\hat{N}$, the lemma follows.

Comment. Suppose in Lemma 2 we assume that N is not orientable but that ∂^A is orientable. Let $\tilde{N}$ be the orientable two-sheeted covering of N. Then $\partial\tilde{N} = \tilde{\partial}^A \cup \tilde{\partial}^B$, where $\tilde{\partial}^A$ covers ∂^A and $\tilde{\partial}^B$ covers ∂^B. Since ∂^A is orientable, $\tilde{\partial}^A$ will consist of two copies of ∂^A and hence will not be connected. Consequently, we can apply Lemma 2 in the orientable case to conclude that there exists a minimal 2-sphere (or projective plane) in $\tilde{N} \backslash \partial\tilde{N}$ which is locally of least area. This observation allows us to rule out nonorientable black holes.

3. The Variational Argument

Let V be a standard static body surrounding k black holes, and suppose V does not have the simple topology described in the main theorem of Sect. 1. Then, in view of the topological lemma, and the comment following its proof, we can assume without loss of generality (by passing to an appropriate covering manifold if necessary) that V contains a minimal 2-sphere which is locally of least area in the sense described in Sect. 2. The following lemma then yields the main theorem in the case the energy condition EC1 holds.

Lemma 3. *Let V be a static body surrounding k black holes. If the energy condition EC1 holds then $V\backslash\partial V$ cannot contain a minimal 2-sphere which is locally of least area.*

Proof. Suppose to the contrary that Σ is a minimal 2-sphere in $V\backslash\partial V$ which is locally of least area. Let N be a smooth unit normal along Σ. We use the conformally related metric, $d\bar{\sigma}^2 = \phi^{-2}d\sigma^2$ to define a variation $u \to \Sigma_u$ of $\Sigma = \Sigma_0$ in V. Let $\bar{E}:(-\varepsilon, \varepsilon)\times\Sigma \to V$ be the normal exponential map of Σ with respect to the metric $d\bar{\sigma}^2$, i.e., $\bar{E}(u, q) = \gamma_q(u)$, where γ_q is the $d\bar{\sigma}^2$-geodesic satisfying $\gamma_q(0) = q$ and $\gamma_q'(0) = \phi(q)N_q$. Choose ε sufficiently small so that $\bar{E}$ is a diffeomorphism onto $U = \bar{E}((-\varepsilon, \varepsilon)\times\Sigma)$. Then for each $u\in(-\varepsilon, \varepsilon)$, define $\Sigma_u = \bar{E}(u, \Sigma) = \{\bar{E}(u, q)\colon q\in\Sigma\}$. Thus, Σ_u is the surface obtained by pushing out along the normal geodesics to Σ in the metric $d\bar{\sigma}^2$ a signed distance u. Although we have used the auxilliary metric $d\bar{\sigma}^2$ to define our variation $u \to \Sigma_u$, all subsequent calculations are in the original induced metric $d\sigma^2$.

Let $B = B_u$ be the second fundamental form of Σ_u; thus for vectors $X, Y\in T_p\Sigma_u$, $B(X, Y) = -\langle \nabla_X N, Y\rangle$, where $\langle\, , \rangle$ is the metric on V and N now denotes the unit normal field to the Σ_u's. A straightforward computation in Gaussian normal coordinates u, x^1, x^2 (with respect to the metric $d\bar{\sigma}^2$) yields the following evolution equation for B,

$$\frac{\partial b_{ij}}{\partial u} = \phi^{-1}R_{i3j3} + \phi b_i{}^m b_{mj} + \phi_{;ij}, \quad 1 \leqq i, j \leqq 2\,, \tag{6}$$

where R is the Riemann curvature tensor of V (sign and notational conventions as in [HE]), $b_{ij} = B(\partial_i, \partial_j)$, and $\phi_{;ij} = \mathrm{Hess}_V\phi(\partial_i, \partial_j)$. Taking the trace of Eq. (6) gives,

$$\frac{\partial H}{\partial u} = \phi\,\mathrm{Ric}_V(N, N) + \phi|B|^2 + \Delta_{\Sigma_u}\phi\,, \tag{7}$$

where $H = H_u = \mathrm{tr}\, B_u$ is the mean curvature of Σ_u.

Let $A(u) =$ area of Σ_u. The formula for the first variation of area gives,

$$A'(u) = -\int_{\Sigma_u}\phi H\, dA = -\int_{\Sigma_u}\phi^2\cdot\frac{H}{\phi}\, dA\,. \tag{8}$$

We obtain an evolution equation for $\frac{H}{\phi}$. The static field equations (i.e., the field equations associated with the metric (1)) are as follows,

$$\mathrm{Ric}_V = \mathrm{Ric}_M + \phi^{-1}\,\mathrm{Hess}_V\phi\,, \tag{9}$$

$$\Delta_V\phi = \mathrm{Ric}_M(e_0, e_0)\phi\,, \tag{10}$$

where $e_0 = \frac{X}{\phi}$. The Laplacians $\Delta_V\phi$ and $\Delta_{\Sigma_u}\phi$ are related by,

$$\Delta_V\phi = \Delta_{\Sigma_u}\phi + \mathrm{Hess}_V\phi(N, N) - \frac{H}{\phi}\frac{\partial\phi}{\partial u}\,. \tag{11}$$

Equations (10) and (11) imply,

$$\Delta_{\Sigma_u}\phi = \phi\,\mathrm{Ric}_M(e_0, e_0) - \mathrm{Hess}_V\,\phi(N, N) + \frac{H}{\phi}\frac{\partial\phi}{\partial u}\,. \tag{12}$$

Substituting (9) and (12) into (7) gives,

$$\frac{\partial}{\partial u}\frac{H}{\phi} = \mathrm{Ric}_M(Z, Z) + |B|^2\,, \tag{13}$$

where Z is the null vector field $e_0 + N$ along V.

Since Σ is minimal, $H_0 = 0$ and $A'(0) = 0$. By EC1 and (13), $H_u \geqq 0$ for $u \in [0, \varepsilon)$, and hence by (8), $A'(u) \leqq 0$ for $u \in [0, \varepsilon)$. If A' ever becomes strictly negative on $(0, \varepsilon)$ then for some $u \in (0, \varepsilon)$, $A(u) < A(0)$, contradicting the fact that Σ is locally of least area. Thus, we must have $A'(u) = 0$ for all $u \in [0, \varepsilon)$, and hence by (8), $H_u = 0$ for $u \in [0, \varepsilon)$. EC1 and (13) now imply,

$$\mathrm{Ric}_M = 0 \quad \text{and} \quad B = 0 \tag{14}$$

on $U \cap \{0 \leqq u < \varepsilon\}$. By using the same variation, but with respect to the normal $-N$, we can conclude, in fact, that Eq. (14) hold on all of U. Equations (6) and (9), taken in conjunction with (14), imply that for all vectors X, $Y \in T_p\Sigma$,

$$\phi^{-1}\,\mathrm{Hess}_\Sigma\,\phi(X, Y) = \mathrm{Ric}_V(X, Y) = -\langle R(X, N)N, Y\rangle\,. \tag{15}$$

We have used the fact that since Σ is totally geodesic, $\mathrm{Hess}_V\,\phi(X, Y) = \mathrm{Hess}_\Sigma\phi(X, Y)$ for vectors X, Y tangent to Σ.

For any $p \in \Sigma$, let $\{e_1, e_2, e_3 = N\}$ be an orthonormal basis of T_pV, and let $K(e_i, e_j)$ denote the sectional curvature in V of the plane spanned by $\{e_i, e_j\}$. Since Σ is totally geodesic, $K(e_1, e_2) = K =$ the Gaussian curvature of Σ at p. This observation and (15) imply

$$\begin{aligned}\mathrm{Ric}_V(e_1, e_1) &= K(e_1, e_2) + K(e_1, e_3)\\ &= K - \mathrm{Ric}_V(e_1, e_1)\,.\end{aligned}$$

Hence, for any unit vector $e_1 \in T_p\Sigma$, $\mathrm{Ric}_V(e_1, e_1) = \frac{1}{2}K$. By polarization it follows that for all vectors X, $Y \in T_p\Sigma$,

$$\mathrm{Ric}_V(X, Y) = \frac{1}{2}K g_\Sigma(X, Y)\,, \tag{16}$$

where g_Σ is the induced metric on Σ. Thus, (15) and (16) show that $\phi|_\Sigma$ must satisfy the following tensor equation on Σ,

$$\mathrm{Hess}_\Sigma\,\phi = \frac{1}{2}K\phi g_\Sigma\,. \tag{17}$$

The remainder of the proof consists of showing that there are no global positive solutions ϕ to Eq. (17). With respect to coordinates $\{x^1, x^2\}$ on Σ, (17) can be written as,

$$\phi_{;ij} = \frac{1}{2}K\phi g_{ij}\,. \tag{18}$$

The Ricci identity,

$$\phi_{;ijk} - \phi_{;ikj} = R_{imkj} g^{mn} \phi_{,n} , \tag{19}$$

where R_{imkj} is the Riemann curvature tensor of Σ,

$$R_{imkj} = K(g_{ik}g_{mj} - g_{ij}g_{mk}) , \tag{20}$$

provides an integrability condition for (18). Indeed, (19) and (20) imply,

$$\phi_{;ijk} - \phi_{;ikj} = K(\phi_{,j}g_{ik} - \phi_{,k}g_{ij}) . \tag{21}$$

On the other hand, covariantly differentiating (14) gives,

$$\phi_{;ijk} - \phi_{;ikj} = \frac{1}{2}[(\phi K)_{,k}g_{ij} - (\phi K)_{,j}g_{ik}] . \tag{22}$$

Equating the right-hand sides of (21) and (22), and contracting both sides of the resulting equation with g^{ij} leads to the equation, $(K\phi^3)_{,k} = 0$, i.e. $\mathrm{grad}_\Sigma(K\phi^3) = 0$. Thus, $K = c\phi^{-3}$ for some constant c. Substituting $K = c\phi^{-3}$ into (17) and contracting we obtain, $\Delta_\Sigma \phi = c\phi^{-2}$. Integrating this equation over Σ shows that $c = 0$, and hence that Σ has vanishing Gaussian curvature. Since Σ is topologically a 2-sphere, this contradicts the Gauss–Bonnet theorem. Thus, there are no positive solutions to (17), which concludes the proof of Lemma 3.

4. Extension to the electrostatic case

The proof of the main theorem in the case EC2 holds is complicated by the fact that Lemma 3, with EC1 replaced by EC2, no longer holds. The following electrovac spacetime is a counterexample.

Example. Let $M = \mathbb{R}^2 \times S^2$ with metric

$$ds^2 = -\cosh^2 \beta r\, dt^2 + dr^2 + \beta^{-2}\, d\Omega^2 ,$$

where $d\Omega^2$ is the metric on the standard unit sphere and β is a positive constant. A straightforward computation shows that this spacetime satisfies the Einstein equation, with $\mathscr{T} = \mathscr{E}$, where $\mathscr{E}$ is the electromagnetic energy-momentum tensor associated with the electrostatic field $\vec{E} = \beta\frac{\partial}{\partial r}$. Let $V = [-1, 1] \times S^2$ with metric $d\sigma^2 = dr^2 + \beta^{-2} d\Omega^2$. Then V is a static body surrounding zero black holes ($\partial^H = \emptyset$) with external boundary $\partial^E = \{-1, 1\} \times S^2$. Since ∂^E is totally geodesic, V is not standard. V fails to be a counterexample to the main theorem only in that ∂^E is not strictly mean convex.

The following lemma takes the place of Lemma 3 in the electrostatic case. It shows that the preceding example is essentially the only counterexample to Lemma 3 with EC1 replaced by EC2.

Lemma 4. *Let V be a static body surrounding k black holes and suppose the energy condition* EC2 *holds. If Σ is a minimal* 2*-sphere in $V \setminus \partial V$ which is locally of least area then there exists a neighborhood U of Σ which is isometric to $((-\varepsilon, \varepsilon) \times S^2$, $dr^2 + \beta^{-2} d\Omega^2)$, where $\beta = |\vec{E}|$ is a positive constant.*

Proof of Lemma 4. Let the notation be as in the first two paragraphs of the proof of Lemma 3. With respect to the coordinates u, $x = (x_1, x_2)$ introduced there, the metric of V in $U \approx (-\varepsilon, \varepsilon) \times \Sigma$ has the form,

$$d\sigma^2 = \phi^2(u, x)du^2 + \sum_{i,j=1}^{2} g_{ij}(u, x)dx^i dx^j \,. \tag{23}$$

The proof of Lemma 4 proceeds just as in the proof of Lemma 3 up to Eqs. (14). These equations now become,

$$\mathrm{Ric}_M(Z, Z) = 0 \quad \text{and} \quad B = 0 \text{ on } U \,, \tag{24}$$

where $Z = e_0 + N$. The vanishing of B implies that the g_{ij}'s are independent of u, $\frac{\partial g_{ij}}{\partial u} = 0$. Thus, we can write,

$$d\sigma^2 = \phi^2(u, x)du^2 + dl^2 \,, \tag{25}$$

where $dl^2 = \sum_{i,j=1}^{2} g_{ij}(x)dx^i dx^j$ is the induce metric on Σ.

Using the Einstein equation and property (1) of EC2, the first equation in (24) implies,

$$\mathscr{E}(Z, Z) = 0 \quad \text{and} \quad \mathscr{M} = 0 \text{ on } U \,. \tag{26}$$

Let $\vec{E}$ be the electric field along V as measured in the static frame. We show that on U, $\vec{E}$ must be a multiple of N. Since this is trivially true at points where $\vec{E}$ vanishes, let $p \in U$ be a point at which $\vec{E}$ is nonzero. Introduce an orthonormal basis $e_0 = \frac{X}{\phi}, e_1, e_2, e_3 = \frac{\vec{E}}{|\vec{E}|}$ of T_pM. With respect to this basis the electromagnetic field tensor $\mathscr{F}$ at p becomes,

$$(F_{ij}) = \begin{pmatrix} 0 & 0 & 0 & -|\vec{E}| \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ |\vec{E}| & 0 & 0 & 0 \end{pmatrix},$$

and thus from Eq. (3) we obtain the following for the components of $\mathscr{E}$ at p,

$$(E_{ij}) = \frac{|\vec{E}|^2}{8\pi} \mathrm{diag}(1, 1, 1, -1) \,. \tag{27}$$

Now express Z with respect to the basis $\{e_i\}$, $Z = e_0 + N = e_0 + \sum_{i=1}^{3} N^i e_i$. The conditions: $\langle Z, Z \rangle = \mathscr{E}(Z, Z) = 0$ then imply that $N_1 = N_2 = 0$. Hence, $N = \pm e_3$, which shows that $\vec{E}$ is proportional to N at p.

Thus, there exists a smooth function β: $U \to \mathbb{R}$ such that $\vec{E} = \beta N$ on U. Since $\mathscr{M} = 0$ on U, $\mathscr{F}$ obeys the free space Maxwell equations, $\operatorname{div} \vec{E} = 0$ and $\operatorname{curl} \vec{E} = 0$ on U. Since $\operatorname{div} N = 0$, $\operatorname{div} \vec{E} = 0$ implies that $\frac{\partial \beta}{\partial u} = 0$ and hence $\beta = \beta(x)$ is independent of u. Since (from (23)) $N = \phi \operatorname{grad} u$, $\operatorname{curl} \vec{E} = 0$ implies $\operatorname{grad} \beta\phi \times N = 0$, and hence $\beta\phi$ is independent of x. Thus, there exists a smooth function λ:$(-\varepsilon, \varepsilon) \to \mathbb{R}$ such that

$$\beta(x)\phi(u, x) = \lambda(u) \text{ for all } (u, x) \in U \,. \tag{28}$$

Equation (28) implies that if β vanishes at some point in U then β and hence $\vec{E}$ vanish identically on U. In this case EC1 holds on U and Lemma 3 implies that Σ cannot be locally of least area, contradicting our assumptions. Thus, without loss of generality, we can assume $\beta = \beta(x)$ and $\lambda = \lambda(u)$ are strictly positive. Making use of Eqs. (25) and (28) and the change of variable,

$$v = \int \lambda(u)du$$

the metric $d\sigma^2$ can be written,

$$d\sigma^2 = \beta(x)^{-2}dv^2 + dl^2. \tag{29}$$

The remainder of the proof consists of showing that β is constant and (Σ, dl^2) has constant curvature $K = \beta^2$. From the Einstein equation and the form of $\mathscr{E}$ (cf., (27)) we observe that the spacetime Ricci tensor obeys,

$$\mathrm{Ric}_M(N, N) = -\beta^2, \quad \text{and} \tag{30}$$

$$\mathrm{Ric}_M(X, Y) = \beta^2 g_\Sigma(X, Y), \quad X, Y \in T_p\Sigma\,. \tag{31}$$

By making use of (31) and modifying the arguments in Sect. 3 in a straightforward way, Eqs. (15), (16) and (17) become for $X, Y \in T_p\Sigma$,

$$\phi^{-1}\mathrm{Hess}_\Sigma\,\phi(X, Y) = \mathrm{Ric}_V(X, Y) - \beta^2 g_\Sigma(X, Y) = -\langle R(X, N)N, Y\rangle\,, \tag{32}$$

$$\mathrm{Ric}_V(X, Y) = \frac{1}{2}(K + \beta^2)g_\Sigma(X, Y), \quad \text{and} \tag{33}$$

$$\mathrm{Hess}_\Sigma\phi = \frac{1}{2}(K - \beta^2)\phi g_\Sigma\,. \tag{34}$$

By suitably modifying the argument in Sect. 3, we obtain the following integrability condition for Eq. (34),

$$\mathrm{grad}_\Sigma(\phi^3 K + \lambda^2\phi) = 0\,,$$

from which it follows that there exists a constant a such that,

$$K = a\beta^3 - \beta^2\,. \tag{35}$$

Equation (7), and the trace of (34) imply,

$$\mathrm{Ric}_V(N, N) = \beta^2 - K\,.$$

Combining the previous equation with (9) and (30) gives,

$$\phi^{-1}\mathrm{Hess}_V\,\phi(N, N) = 2\beta^2 - K\,. \tag{36}$$

On the other hand, noting that $N = \beta\dfrac{\partial}{\partial v}$, one calculates,

$$\phi^{-1}\mathrm{Hess}_V\,\phi(N, N) = N(N(\phi)) - \nabla_N N(\phi) = \beta^2\frac{1}{\lambda}\frac{\partial^2\lambda}{\partial v^2} - \frac{|\nabla_\Sigma\beta|^2}{\beta^2}, \tag{37}$$

where $\nabla_\Sigma = \mathrm{grad}_\Sigma$.

By equating the right-hand sides of (36) and (37) and using (35) we obtain,

$$\frac{|\nabla_\Sigma\beta|^2}{\beta^4} - a\beta = \text{constant} \tag{38}$$

on Σ. We now differentiate both sides of (38) along Σ; with respect to an arbitrary vector $X \in T_p\Sigma$ we have,

$$\begin{aligned} X\left(\frac{|\nabla_\Sigma \beta|^2}{\beta^4}\right) &= X\left\langle \nabla_\Sigma \frac{1}{\beta}, \nabla_\Sigma \frac{1}{\beta}\right\rangle = -\frac{2}{\beta^2}\left\langle \nabla_X \nabla_\Sigma \frac{1}{\beta}, \nabla_\Sigma \beta\right\rangle \\ &= -\frac{2}{\beta^2}\mathrm{Hess}_\Sigma \frac{1}{\beta}(X, \nabla_\Sigma \beta) \\ &= -\frac{1}{\beta^3}(K - \beta^2) g_\Sigma(X, \nabla_\Sigma \beta) \quad \text{(cf. Eq. (34))} \\ &= \left(\frac{2}{\beta} - a\right) X(\beta) . \end{aligned} \tag{39}$$

Thus, from (38) and (39) we arrive at,

$$\left(\frac{1}{\beta} - a\right) \nabla_\Sigma \beta = 0 ,$$

which implies that β is constant. Hence, the left-hand side of (34) is zero, and so $K = \beta^2$. Thus, $dl^2 = \beta^{-2} d\Omega^2$, and (29) becomes,

$$d\sigma^2 = dr^2 + \beta^{-2} d\Omega^2 ,$$

where $r = \frac{v}{\beta}$. This concludes the proof of Lemma 4.

Finally, we indicate how Lemma 4 is used to prove the main theorem in the case that EC2 holds. Suppose V does not have the desired simple topology. Then, by the topological lemma in Sect. 2, we can assume V contains a minimal 2-sphere Σ which is locally of least area. By Lemma 4, Σ is a totally geodesic round sphere, and a neighborhood U of Σ is isometric to a metric product $(-\varepsilon, \varepsilon) \times \Sigma$. Let γ be a shortest geodesic from Σ to ∂^E, and let Σ' be the component of ∂U which meets γ. It is easy to see that the product structure of U extends to Σ' and hence that Σ' is metrically parallel to Σ. Moreover, since Σ arises from minimizing area in isotopy class (i.e., Σ equals $\Sigma^{(j)}$ for some j), the area minimizing properties of the $\Sigma^{(j)}$'s guarantee that Σ' is locally of least area. Lemma 4 applied to Σ' implies that the product structure of U can be extended beyond Σ'. Hence, by a straightforward continuation argument, the product structure of U extends all the way along γ out to a component of ∂^E, thereby forcing this component to be totally geodesic. But this contradicts the assumption that ∂^E is strictly mean convex.

Concluding remarks. In this paper we have shown from local hypotheses that, provided appropriate energy conditions are satisfied, the steady state topology of space (i.e. the topology of space in the static limit) in the vicinity of black holes must be simple. Even in the no black hole case ($\partial^H = \emptyset$), our main theorem improves certain aspects of results obtained in [FG] and [M-A]. The next logical step in this approach to studying black holes is to extend the present work to the stationary (rotating black hole) case. It may be possible to refine the technique used in [FG] to study the topology of stationary fluid bodies to make it applicable to this situation.

Of course, if the energy conditions are violated, it is possible to have equilibrium states with nontrivial spatial topology, such as worm holes. Whether certain quantum mechanical processes can produce the necessary violations has been considered, for example, in [MTY].

Finally, our main theorem has some bearing on the classical black hole uniqueness theorems (cf. [BM] and references cited therein). These theorems require the assumption of asymptotic flatness. In order to express the fall-off of the metric in precise terms, one imposes a priori the existence of Euclidean coordinates at infinity. Using the topological fact that a manifold which can be expressed as the increasing union of open balls is diffeomorphic to $\mathbb{R}^3$, our main theorem provides mild geometric conditions for the existence of such coordinates.

References

[BM] Bunting, G.L., Masood-ul-Alam, A.K.M.: Nonexistence of multiple black holes in asymptotically euclidean static vacuum space-times. Gen. Rel. Grav. J. **19**, 147–154 (1987)

[FG] Frankel, T., Galloway, G.J.: Stable minimal surfaces and spatial topology in general relativity. Math. Zeit. **181**, 396–406 (1982)

[E] Eschenburg, J.-H.: Maximum principle for hypersurfaces. Manuscripta Math. **64**, 55–75 (1989)

[G] Gannon, D.: Singularities in nonsimply connected space-times. J. Math. Phys. **16**, 2364–2367 (1975)

[HE] Hawking, S.W., Ellis, G.F.R.: The large scale structure of space-time. Cambridge: Cambridge University Press 1973

[L] Lee, C.W.: A restriction on the topology of Cauchy surfaces in general relativity. Commun. Math. Phys. **51**, 157–162 (1976)

[M-A] Masood-ul-Alam, A.K.M.: The topology of asymptotically euclidean static perfect fluid space-time. Commun. Math. Phys. **108**, 193–211 (1987)

[MSY] Meeks, W., Simon, L., Yau, S.-T.: Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature. Ann. Math. **116**, 621–659 (1982)

[MTY] Morris, M.S., Thorne, K.S., Yurtsever, U.: Wormholes, time machines, and the weak energy condition. Phys. Rev. Lett. **61**, 1446–1449 (1988)

Communicated by S.-T. Yau