Lost Horizon? – Modeling Black Holes in String Theory

Nick Huggett & Keizo Matsubara[*]

Department of Philosophy
University of Illinois at Chicago
601 South Morgan Street
Chicago, Illinois 60607, USA
huggett@uic.edu

Department of Philosophy
Uppsala University
Box 627, SE-751 26 Uppsala, Sweden
keizo.matsubara@filosofi.uu.se

Abstract

The modeling of black holes is an important desideratum for any quantum theory of gravity. Not only is a classical black hole metric sought, but also agreement with the laws of black hole thermodynamics. In this paper, we describe how these goals are obtained in string theory. We review black hole thermodynamics, and then explicate the general stringy derivation of classical spacetimes, the construction of a simple black hole solution, and the derivation of its entropy. With that in hand, we address some important philosophical and conceptual questions: the confirmatory value of the derivation, the bearing of the model on recent discussions of the so-called 'information paradox', and the implications of the model for the nature of space.

[*]Some of this work was performed under a collaborative agreement between the University of Illinois at Chicago and the University of Geneva and made possible by grant numbers 56314 and 61387 from the John Templeton Foundation. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the John Templeton Foundation. We also thank John Dougherty for considerable assistance at the start of this work.

1 Introduction

In their article on singularities and black holes in the Stanford Encyclopedia of Philosophy, Peter Bokulich and Erik Curiel raise a series of important philosophical questions regarding black holes, including the following:

> When matter forms a black hole, it is transformed into a purely gravitational entity. When a black hole evaporates, spacetime curvature is transformed into ordinary matter. Thus black holes appear to be crucial for our understanding of the relationship between matter and spacetime, and so provide an important arena for investigating the ontology of spacetime, of material systems, and of the relations between them. Curiel and Bokulich [2012]

This paper develops this insight to investigate the natures and relations of spacetime and matter in quantum gravity, specifically in string theory. Part of the paper will therefore be devoted to explicating the general status of spacetime in string theory (§3), and especially its emergence, and fungibility with matter; and to outlining a well-studied example of a string theoretic black hole (§4).

As Bokulich and Curiel note, of particular significance in such an investigation is the phenomenon of black hole thermodynamics (BHT) and Hawking radiation. This topic has been widely discussed by philosophers of physics as well as physicists, so we will just give a brief review (§2). It is important to note that while much theoretical work motivates the results of BHT, there is no direct empirical confirmation of these results. The only experimental evidence comes from work on analogue systems, whose significance remains controversial.[1] The result that black holes radiate is, nonetheless, generally trusted since the derivations rely on well tested theories, applied in regimes where we should be able to trust the derived conclusions. Once we have described a stringy black hole, and the agreement of Boltzmann and Bekenstein-Hawking entropies, we will discuss the epistemic significance of this result (§5.1).

However, our focus is on the ontological aspects of black holes, and we will argue (§5) that the issues that arise under the heading of the 'information paradox', such as the unitarity of black hole evaporation, and the possibility of 'firewalls' or 'fuzzballs', suggest insights into the nature of spacetime in

[1] For instance, see Dardashti et al. [2017], and Crowther et al. [2019].

the interior of a stringy black hole. Then we will turn to some more general lessons about the relation between space and matter, to be drawn from our discussion (§6).

2 Black hole thermodynamics

Assuming some familiarity concerning the topic of BHT, this review will be brief.[2] Starting with classical general relativity (GR) without quantum effects, it was observed that black holes seemed to violate the second law of thermodynamics: dropping things into a black hole could seemingly destroy entropy. To avoid this conclusion Bekenstein [1973] proposed an entropy proportional to the area of the horizon. Keeping all physical constants explicit, the formula for the entropy is as follows:

$$S_{BH} = \frac{k_B c^3 A}{4\hbar G} = \frac{k_B A}{4\ell_p^2} \tag{1}$$

using that the Planck length $\ell_p = \sqrt{\frac{G\hbar}{c^3}}$, to display more explicitly how the area of the black hole is divided into Planck length squared areas.

However, there is an incompatibility between the Bekenstein entropy, the Boltzmannian understanding of thermodynamics, and the "no hair theorem". On the one hand the entropy should be attributed to the (logarithm of) the number of black hole microstates. On the other, in classical GR the state of a black hole is *completely* characterized by its mass, charge and angular momentum – "black holes have no hair". So classical black holes simply don't have the microstates necessary to understand the Bekenstein entropy in Boltzmannian terms.

Perhaps black holes have a novel, non-Boltzmannian form of entropy; this was the view originally endorsed by Hawking [1976, 1975]. However, physicists working in the different approaches to quantum gravity generally aim to provide a description of the quantum microphysics of black holes. If such an account can be given, the Boltzmann picture "assures" us that some form of the second law holds even when systems include black holes: by state

[2] That black holes have entropy was originally made in Bekenstein [1973]. After Hawking [1975] showed that black holes radiate, BHT was taken much more seriously. Philosophical work on BHT include Belot et al. [1999], Wallace [2018, 2019, 2020], and Wüthrich [2017]. Reviews by physicists include Susskind and Lindesay [2005], Mathur [2009], Harlow [2016], and Polchinski [2017].

space volume considerations, most states at lower entropy evolve to states of higher entropy.

Now, if black holes are properly thermodynamical, then there should also be a temperature associated with them, and they should seek thermal equilibrium with their environment. Of course, Hawking radiation provides a realization of just this. However, it also allows for ‘information loss’: in spacetimes containing black holes, pure quantum states can evolve into mixed states. However, quantum physics is unitary, and it is a theorem that unitarity prohibits the evolution from a pure state to mixed one. Moreover, such an evolution amounts to a failure of *backwards* determinism: one cannot retrodict an earlier pure state from a mixed state. This surprising conclusion led to much debate on the so-called “black hole information paradox” (or “problem”).

One response is ‘black hole complementarity’[3], whose central idea is that external observers never see matter entering the horizon, because of the infinite red shift, and instead observe it radiating back in an unproblematic way. Observers that do cross the horizon of course do see matter entering the black hole, but are shielded from observing any inconsistency (specifically, violations of the quantum no cloning theorem) because they fall into the singularity too quickly. In response, Almheiri et al. [2013] aimed to show that three claims assumed by complementarity are inconsistent.[4] Quote:

1. **Unitarity**: Hawking radiation is in a pure state.
2. **Semi-classical gravity**: The information carried by the radiation is emitted near the horizon, with low energy effective field theory valid beyond some distance from the horizon.
3. **No drama**: The infalling observer encounters nothing special at the horizon.

To avoid contradiction, Almheiri et al. deny (3), proposing that an infalling observer does not pass the horizon as expected classically, but instead is destroyed by a ‘firewall’; which certainly would be drama! In §5.2.2 we will return briefly to firewalls; while in §5.2.3 we shall see another view, which also rejects 3.

[3] See Susskind et al. [1993].

[4] Black hole complementarity has also been attacked on philosophical grounds by Belot et al. [1999].

3 Spacetime in string theory

Before continuing with these issues, we need to outline sufficient string theory to understand how stringy black holes arise: the origin of both spacetime (§3.1) and matter (§3.2) in the theory.

3.1 Spacetime in string theory: fungibility of geometry and matter

Huggett and Vistarini [2015] explained the derivation of the Einstein Field Equation (EFE) – the 'emergence' of GR – in string theory. Since this story is central to the points of this paper we must review it, but with emphasis on the conceptual picture, and without the technical details found in that paper.[5]

The starting point for classical string theory is the Nambu-Goto action, which tells us to extremize the worldsheet spacetime area of a string in a d-dimensional Minkowski background (figure 1). So doing leads to a relativistic wave equation, with either Neumann (momentum conserving) or Dirichlet (position conserving) boundary conditions at the end points.

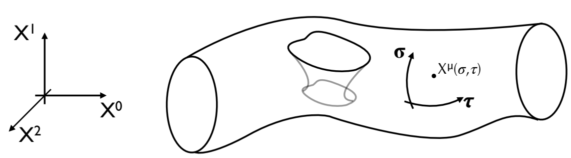


Fig. 1: A closed string in spacetime. The trajectory is described by an embedding function from worldsheet coordinates to spacetime coordinates: $(\sigma, \tau) \to X^\mu$.

However, the Nambu-Goto formulation is infelicitous for quantization, so one shifts to the classically equivalent Polyakov action (see (2) below). So doing introduces an 'auxiliary' Lorentzian metric $h_{\alpha\beta}$ on the string worldsheet, *distinct* from the metric 'induced' on the world sheet by the Minkowski metric of background spacetime. (The subscripts range over the two coordinates σ and τ on the worldsheet.) Importantly, the action has 'Weyl symmetry' with respect to $h_{\alpha\beta}$: $h_{\alpha\beta} \to e^{\Omega(\sigma,\tau)} h_{\alpha\beta}$ for any smooth real function $\Omega(\sigma, \tau)$.

[5] Or in the sources from which it is drawn, e.g. Polchinski [1998]. See Vistarini [2019] or Huggett and Wüthrich [forthcoming] for longer philosophical analyses.

Thus there is no physical significance to the auxiliary metric beyond the causal structure it ascribes to the string, which must agree with that of the background spacetime in order to minimize the action.

On canonical quantization, the classical wave solutions become quanta *on* the string, in the way familiar from quantum field theory (QFT), which when grouped into states of equal energy form representations of $SO(1, d-1)$, just like relativistic particles in d-dimensional spacetime. Hence particles are reinterpreted as strings in the appropriate representations, with rest mass associated with the vibrational energy of the string – at length scales at which the string is indistinguishable from a point. By this mechanism string theory promises to unify the different fundamental particles: they are nothing but different modes of a single underlying object, the string, and hence fungible if the state of the string changes. In particular, the spectrum of the closed bosonic string contains the massless spin-2 representation that characterizes the graviton, the quantum of the metric field; these modes/particles are therefore in particular fungible with those of other fields. That said, several points should be made.

First, we are yet to identify *quanta* of the corresponding quantum fields as strings, since creation and annihilation of quanta requires creation and annihilation of strings, about which nothing has yet been said. Modes on a string can be created and annihilated, but that does not change the number of strings, just the kind of particle that a string represents. Second, massless spin-2 fields lead almost inevitably to GR: classically see Misner and Thorne [1973, §18.1], while Salimkhani [2018] reviews the situation in QFT. So if this mode of the string truly is a quantum of the gravitational field, we need to verify that it relates dynamically to other fields in the appropriate way – through the EFE. Third, the bosonic string is incapable of reproducing the mass spectrum of the standard model; again, more structure must be added. All three points will be developed later.

Progressing further requires shifting to a path integral approach, in which each path contributes an amplitude equal to the exponential of its action, or rather e^{iS}. Wick rotating the worldsheet coordinates $\tau \to i\tau$ to give the auxiliary metric $h_{\alpha\beta}$ a Euclidean signature, the Polyakov path integral is given by Polchinski [1998, §3.2]:

$$\int_{paths} DXDh\ \exp\left\{\frac{-1}{4\pi\alpha'}\int_M \mathrm{d}\sigma\mathrm{d}\tau\ h^{1/2}h^{\alpha\beta}g_{\mu\nu}\partial_\alpha X^\mu\partial_\beta X^\nu\right\}, \tag{2}$$

where the 'Regge slope' α' is the characteristic string length squared, M is a specified worldsheet, and (for now) $g_{\mu\nu} = \eta_{\mu\nu}$, a background Minkowski metric. The path integral is taken over all embeddings X^{μ} and all auxiliary metrics $h_{\alpha\beta}$.

The path integral involves a sum over all topologically distinct worldsheets: for the closed string, tori of all possible genera, with N open holes representing in/out strings at temporal infinity. The topological holes in the tori are produced by strings splitting/joining: for instance, figure 1 is a simple torus with $N = 2$, representing a single incoming closed string splitting into two strings, which then recombine into a single outgoing string. The tori therefore represent a perturbative sum of Feynman diagrams, in analogy with those for QFT (indeed under the identification of quanta with string modes, QFT diagrams are understood as approximations to stringy diagrams).

Therefore they assume the existence of a theory in which strings can be created and annihilated, or at least a theory in which Fock-like string states are a reasonable approximation (in some sector). (In)famously, this theory – 'M-theory' – is not known, and so string theory as we are discussing it is inherently perturbative.[6] However, once one accepts this perturbative understanding then the identification of strings with the quanta of QFT is complete: any field state (in the Fock representation, a superposition of different numbers of quanta) is fundamentally a state of many strings (a superposition of different numbers of strings, each in the mode corresponding to the quantum of the field). Thus all fields are unified, composed of strings, differing only in their modes, and fungible if the strings change mode. We now have all the conceptual ingredients needed to understand the origin of GR in string theory.

(i) First, GR allows for curved background spacetime metrics, not just Minkowski spacetime. In QFT, classical fields are represented by 'coherent states' of field quanta. Such states can be defined in various ways (see [Duncan, 2012, §8.2-3]), but two conceptions are salient: first, they are maximally classical, simultaneously minimizing the uncertainty in the canonical variables; second, they are collective states, involving a superposition of every number of field quantum (and so are not finite superpositions). But if a classical field is described by a coherent state of quanta, then according to the identification of quanta with strings, a classical field should correspond to a suitable collective superposition of strings, each excited into the same

[6] A bosonic string field theory, with a 3-point interaction exists (e.g., Taylor [2009]), but is no longer viewed as a promising approach to M-theory.

mode. The story will be the same for any classical field, including a metric field comprised of stringy gravitons.

One can check this identification, by inserting classical fields into the Polyakov action, and comparing the effect on scattering amplitudes with that of scattering in the presence of the corresponding collective string states. For instance, one might take $g_{\mu\nu}$ to be a general spacetime metric rather than Minkowski, and compare it with scattering in a background of a suitable coherent state of stringy gravitons. The results are exactly the same: these are equivalent descriptions.[7] Note that the classical fields are called 'background' fields, but in the sense that they describe a fully stringy background, not because they are added to the theory from the outside.

(ii) Second, a path integral like (2) with a general curved metric is known as a 'non-linear sigma model'; broadly, it describes a field X^μ living on a 2-dimensional spacetime (the string worldsheet) with *variable* interaction $g_{\mu\nu}(X^\mu)$. The crucial result for our purposes is that this quantum theory will only retain the Weyl invariance of the classical action – as it must do in order to avoid a pathological 'anomaly' – if the background metric $g_{\mu\nu}$ and any other background fields satisfy the EFE (to lowest order in α'). For (2), in which there is only a background metric field, the result is the free field equation $R = 0$; in general, with additional background fields, the *full* non-linear equation is entailed.[8] Of course, from our previous discussion, we recognize that the metric (and other) fields are in fact nothing but collective string states.

To summarize: avoiding a Weyl anomaly requires that background fields, including the metric, satisfy the EFE to lowest order in perturbation theory. Physically however, the background does not comprise classical fields in a classical spacetime: rather strings in appropriate modes form coherent states of effective QFTs, which in turn form effective classical fields. So ultimately the Weyl anomaly is a constraint on multi-string states, and the ontology of

[7] Green et al. [1987, §3.4.1] is the earliest presentation of this point of which we are aware. The idea is that a coherent state of strings, each in a massless spin-2 state, introduces a term $\gamma_{\mu\nu}$ in the path integral (2) which adds to the Minkowski metric to produce $\eta_{\mu\nu} \to g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu}$. Since the path integral determines all physical quantities in a quantum theory, we have fully equivalent theories whether we introduce the curved metric as a classical field or as a graviton state.

[8] It is worth stressing that the expansion is in α', so that the approximation is prima facie valid when the radius of spacetime curvature is small compared to the string length: say, compared to the Planck length – far beyond the regime of linear gravity. We will, however, see that it does break down in a 'fuzzball', even for moderate curvature.

fields is one of strings only. But since the quanta of different fields, including the metric, are nothing but different string modes, they are fungible, so that gravity is on the same footing with any other force.[9]

3.2 Supergravity: stringy fermions, gauge fields, and p-branes

Since the world contains fermions one must extend string theory: as bosons arise from spatial modes, so fermions arise from vibrations in 'anti-commuting directions'. A full discussion is well beyond the scope of this paper so we will only sketch points necessary for our string theoretic black hole model. The most important point is that the recovery of GR from string theory just described applies *mutatis mutandis* to superstring theory.

In very general terms, 'supersymmetric' (SUSY) string theory is developed as for the bosonic string. First introduce an action that adds fermionic degrees of freedom $\psi^\mu(\sigma, \tau,)$ (a Majorana spinor) to the bosonic ones $X^\mu(\sigma, \tau,)$:

$$\int_{paths} DXDh\ \exp\big\{\frac{-1}{4\pi\alpha'}\int_M \mathrm{d}\sigma\mathrm{d}\tau\ h^{1/2}h^{\alpha\beta}g_{\mu\nu}(\partial_\alpha X^\mu\partial_\beta X^\nu - i\psi^{\dagger\mu}\rho_\alpha\partial_\beta\psi^\nu)\big\}, \tag{3}$$

where ρ^α are worldsheet Dirac matrices. Green et al. [1987, §4.1] discusses this action, and shows that it possesses classical supersymmetry. There are new endpoint boundary conditions for the fermionic degrees of freedom – not Neumann and Dirichlet, but Ramond or Neveu-Schwarz – and correspondingly new modes. When one canonically quantizes, one's choice of boundary conditions produces a particular spectrum of bosons and fermions. Because of the underlying SUSY these are paired (in addition to Green et al. [1987, §4.2], Zwiebach [2004, chapters 14-6] contains an approachable introduction): each mode is fungible with its 'superpartner', under a symmetry of the theory.

Proceeding exactly as before, the bosonic modes correspond to field quanta, but now of gauge fields. Coherent states of strings in the same mode thus have effective descriptions as classical gauge potentials, $A_\mu, A_{\mu\nu}, A_{\lambda\mu\nu}$, and so on. And of course to avoid the Weyl anomaly, with the metric these mutually

[9] True, the full metric contains Minkowski and stringy parts: $g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu}$. But the conclusion that $\eta_{\mu\nu}$ is a non-stringy classical background can be resisted: Witten [1996], Matsubara [2013], Huggett [2015], Motl [2012]. See Read [2019] for more on fungibility.

satisfy the appropriate EFE, and hence because of their supersymmetry form models of classical 'supergravity'.

The question arises of the sources of these fields. $(n-1)$-dimensional bodies can couple 'electrically' to an n-form gauge field. For instance, a 0-dimensional point body couples as $A_\mu \frac{\mathrm{d}x^\mu(\tau)}{\mathrm{d}\tau}$: the dimension of the body determines whether it has enough indices to 'eat' the field indices. Similarly, $d-n-3$ dimensional objects couple 'magnetically' (since they have to 'eat' the indices on the field's Hodge dual). So the presence of gauge fields speaks for the presence of charged multidimensional objects, known as 'p-branes'. A discussion of their nature in the conceptual framework laid out here will have to wait for another occasion (they are typically thought of in terms of some stable 'solitonic' multi-string state). For now note that they also ground Dirichlet boundary conditions in string theory: if the end of a string is constrained to move within a p-brane, then it is fixed with respect to the remaining $d-1-p$ spatial dimensions. A p-brane to which open strings can attach is thus known as a Dp-brane.

Pulling this together, one of the choices of boundary condition leads to 'type IIB' superstring theory, which contains a 2-form gauge field $B_{\mu\nu}$. So, for example in 10 spacetime dimensions, D1-branes couple electrically and D5-branes magnetically to $B_{\mu\nu}$, and so may be present in a supergravity limit of type IIB superstring theory. In our model, a construction of these branes produces the black hole.

4 A stringy black hole

In this section we sketch a realization of these ideas, a stringy black hole, which will be the basis for the following discussion. Our model is physically unrealistic (at least for our universe), but it is simple yet exhibits the principles behind more realistic examples (hence it is popular in pedagogical presentations, e.g. Das and Mathur [2000] and Zwiebach [2004, chapter 22]). The origin of this type of construction is Strominger and Vafa [1996]. The specific approach discussed was proposed in Horowitz et al. [1996], which they used to show that the entropy agrees with the Bekenstein-Hawking formula (1).

We work in type IIB theory with its D1- and D5-branes, and suppose a background spacetime topology of $R^5 \times S^1 \times T^4$ with coordinates $(x_0, \ldots x_4, x_5, x_6, \ldots x_9)$, respectively. We are interested in the black hole appearing in the

5-dimensional spacetime described by $(x_0, \ldots x_4)$ with topology R^5, and stipulate that the remaining compact dimensions are 'small'. However, the circumference C of the circular S^1 x_5 dimension is much larger than that of the toroidal T^4 $(x_6, \ldots x_9)$ dimensions. The effect of this stipulation is that the minimum wavelength on the torus is much shorter than on the circle, so that the energy cost of excitations on the torus is much greater, and effectively any momentum in the compactified dimensions will be on the circle. Thus if N is the wavenumber on S^1, then C/N is the wavelength, and the internal momentum of the black hole is $P = hN/C$.

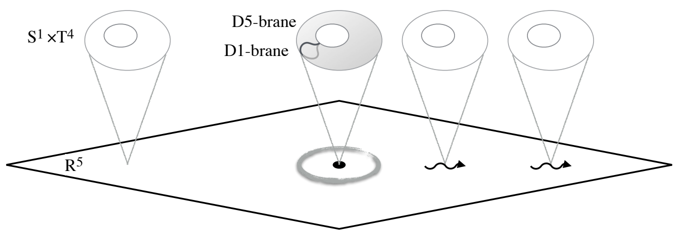


Fig. 2: A stringy black hole: the background spacetime has a topology $R^5 \times S^1 \times T^4$ – time is not shown, and of space S^1, two dimensions of R^4 and one of T^4 are pictured. At a point of R^4 are located D1-branes around S^1 and D5-branes around $S^1 \times T^4$. If the string interaction is 'turned on', a spatial horizon forms around the branes in R^5, and gravitons are radiated.

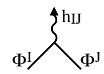
At the origin of the uncompactified space, $(x_1, \ldots x_4) = (0, 0, 0, 0)$, are located (a) Q_1 D1-branes wrapped around S^1, (b) Q_5 D5-branes wrapped around $S^1 \times T^4$, and (c) momentum P (in the x_5 direction, as discussed); see figure 2. These are the sources of the gravitational field (not the field itself). As we saw, the Dp-branes couple to the $B_{\mu\nu}$ gauge field of the theory (whose stringy nature we again emphasize), while P is a source for the metric field $g_{\mu\nu}$ (likewise). Because the EFE holds for such background fields (to avoid the Weyl anomaly) the spacetime geometry in which the system lives can be computed, yielding a model of supergravity possessing a horizon in the four spatial dimensions of R^5, around the origin, as shown in figure 2.

The next step is to apply the technique of 'dimensional reduction' based on the work of Kaluza and Klein (see Karaca [2012]), used in string theory to determine the projection of higher dimensional physics into the large dimensions that we directly observe: gauge fields project into gauge fields. But so does the metric: from the point of view of the large dimensions, the geometry of the compact dimensions acts as if there was a new gauge field – the basis of the Kaluza-Klein scheme to 'geometrize' gauge fields. The upshot in our model is that the R^5 description of the solution is a Reissner-Nordström black hole with three point charges Q_1, Q_5, and P, and mass equal to its internal energy, located at the origin.

The point of constructing such models was to compare their Boltzmann entropy, calculated by counting the number of microstates of such an assembly of branes, with S_{BH} (1), calculated for the dimensionally reduced supergravity black hole. The calculation is described in the references given, but the significance of the models is that these entropies agree: $S \sim \sqrt{Q_1 Q_2 P}$. We will discuss the significance of this result in §5.1, but it should be noted however that the key element in this result is that the system is in a Bogom ol'nyi-Prasad-Sommerfield (BPS) state of superstrings. These arise in SUSY because of the special symmetries (Das and Mathur [2000, §3.2] gives a simple illustration), but have the features that (a) they are energetically stable because of a selection rule, and (b) varying potential terms does not cause any splitting of energy levels. Because of (b) the number of microstates would be the same if the strings were non-interacting, a scenario in which the number of states is understood and computable: the Boltzmann entropy is the same when the interaction is 'turned on'. But because of (a) the black hole in the effective supergravity model is 'extremal', unable to Hawking radiate any further, though not completely evaporated away.[10]

However, as Wadia [2001] explains, one can perturbatively model a near-extremal black hole, and verify that its Boltzmann and Bekenstein-Hawking entropies agree as well. Most significant for our discussion, there is a channel by which branes can radiate gravitons into R^5. That is, if Φ^I ($I = 7, 8, 9, 10$) represents a quantum of D1-brane vibration in the T^4 directions, and h_{IJ} a graviton polarized in the T^4 dimensions propagating in R^5, then the following interaction exists:

[10] An earlier program due to Susskind, on which he reflects in Susskind [2006], approached the same problem by adiabaticity; that slowly lowering the string interaction to zero would not change the state counting. This method is more general, allowing the Boltzmann entropy to be calculated for a range of realistic, non-extremal black holes, but is less reliable because it doesn't have the BPS guarantee that the density of states is constant.



(4)

That is, the model has a mechanism for the black hole to radiate mass away; moreover, the energy cross-section of this radiation agrees with that computed semi-classically for Hawking radiation. Such an interaction thus provides a specific instance of how the fungibility of string modes, especially those of matter and geometry, play out in dynamical processes.

5 String theory and black hole thermodynamics

So far we have described, with an eye to conceptual significance, black hole thermodynamics, and the string theoretic understanding of spacetime and black holes. In the following we turn to an investigation of their philosophical consequences.

5.1 Significance of the derivation

While the focus of this paper is the ontology of stringy black holes, some brief comments on their epistemic import are in order. Especially, what is the confirmatory value for string theory of the equality of Boltzmann and Bekenstein-Hawking entropies? Why are such results considered important, given that there is no direct empirical confirmation of BHT? We will make four points.[11]

First, there are nonetheless reasons to trust BHT, especially the consilience of many routes to their derivation, across multiple contexts; and with the general framework of thermodynamics, beyond gravitational physics. (And perhaps analogue experiment.)

Second, Bekenstein's discovery was 'surprising', which might make it seem a particularly strong piece of evidence. However, the surprise is not of the evidentially relevant kind. We must distinguish the *anticipation* that P is true from the *probability* that P is true conditional on our background beliefs. For the confirmation of string theory our background beliefs include semi-classical GR, the theory of quanta propagating in curved spacetimes; as we saw, recovering this theory is already part of the support for string theory. But as Hawking showed, BHT is a consequence of semi-classical GR, and

[11] See Wallace [2018, 2019] and van Dongen et al. [2020] for more detailed discussions.

so not independent evidence; the surprise was only that of discovering an unknown logical consequence.

Third, indeed, since string theory is believed to have QFT and GR as effective limits, it 'must' entail BHT. Thus it is better to take the successful entropy derivation for stringy black holes as a consistency check rather than new evidence. For instance, Horowitz et al. [1996, p1] seems to express this attitude. However, success is non-trivial: if one could show the failure of only one model to be consistent with the results of BHT then this would be highly problematic for string theory.

Finally, the derivation provides a concrete account of the microstates of a black hole, showing the validity of the string theoretic principles assumed in modelling it: i.e., the assumptions of the previous section and subsection. These details do go beyond general semi-classical GR, and so do receive confirmation from the derivation of BHT. Moreover, their successful application provides two other kinds of support for string theory. In the first place, the derivation of a Boltzmann entropy is a kind of explanation of the Bekenstein-Hawking entropy (and more speculatively of Hawking radiation); and so provides whatever theoretical support successful explanations give.[12] Second, the constructions licensed by these principles allow the application of string theory to further physical situations, in spacetime and particle physics, increasing its fruitfulness, an important non-empirical virtue of theories.

5.2 Reflections on the black hole information paradox

A common (but not universal) view among physicists – in particular particle physicists and string theorists – is that unitarity must not be violated and that the information loss originally proposed by Hawking is incorrect once a full treatment of quantum gravity is taken into account. It is thus of interest to get a better understanding of what happens when black holes evaporate according to various theories of quantum gravity; in this paper, in string theory, and in particular in the "fuzzball" proposal. Before coming to that we need to make a short detour and address a recent argument against any 'paradox', and the idea of "firewalls".

[12] Note that the derivation undermines the idea mentioned earlier that black holes might have non-Boltzmannian entropy.

5.2.1 On Maudlin's recent critique of the information paradox

Maudlin [2017] recently argued that the whole idea of a paradox is due to a simple mistake. He first observes – reiterating Wald [1994] – that the spacelike surfaces after the evaporation of the black hole are not Cauchy-surfaces: causal curves from the past can end up in the singularity, and fail to reach post-evaporation hypersurfaces. But it is only for an evolution of a pure state from one Cauchy-surface to another that the rules of quantum mechanics imply that the state must remain pure. The more novel suggestion made by Maudlin is that therefore the final mixed state does *not* require that the evolution is not unitary. To make this point he uses a slightly unconventional foliation of spacetime, where some of the Cauchy-surfaces are disconnected; with respect to this foliation the full evolution is unitary.

While we do not dispute these technical claims, we believe that physicists working on the black hole information paradox generally are aware of Wald's argument, and won't be moved by Maudlin's conclusions. In particular, his description of the situation presupposes the classical, GR description of spacetime everywhere, but this cannot be taken for granted in a theory of quantum gravity.[13] It is true that he offers a way to reconcile classical spacetime with unitarity, but it only diagnoses the loss of information rather than removing it. Many working in the field expect a full quantum gravity description of the formation and evaporation of black holes not to involve any singularities or loss of causality, and yet remain unitary. The 'paradox' is that this does not occur in the combination of the two theories, GR and QFT, that are presumably low energy limits of the fundamental theory.[14] Thus from this point on we do assume both unitarity and that external observers do not encounter information loss.

5.2.2 Firewalls

In §2 we described the 'AMPS' argument (Almheiri et al. [2013]) that the premises assumed by black hole complementarity were not consistent. A number of different responses have been formulated (see Polchinski [2017] and Harlow [2016]). One is to accept 'drama' at the horizon, or even the absence of a horizon in the first place. Objects – and observers – never really

[13] Huggett and Wüthrich [2013] explores spacetime emergence.

[14] In addition, see Wallace [2020] for a convincing demonstration that there are forms of the paradox that resist Maudlin's analysis.

pass the horizon, instead they are thermalized before they reach it; there only is the exterior description, no complementary description according to an infalling observer. In this case, it has been suggested that there is no classical spacetime interior either:

> Finally, since we are thinking that spacetime is emergent, we might try the slogan that it is not that the firewall appears, but that the interior spacetime fails to emerge. But to claim this we would need a better understanding of emergent spacetime. (Polchinski [2017, 31].)[15]

From Maudlin's point of view this suggestion might seem irrelevant, since he does not accept the premises that motivated the introduction of the firewall in the first place. But from the point of view of a quantum theory of gravity in which spacetime is emergent, such a view certainly makes sense; the interior could be described by fundamental degrees of freedom that do not have a classical spacetime description. And if the black hole interior is eliminated, then the disconnected parts of the Cauchy-slices to which Maudlin appeals for unitarity will also be eliminated. As we noted, we believe that Maudlin's reasoning takes a fundamental spacetime for granted, when this is often denied by those in the debate.

5.2.3 Fuzzballs

The idea of firewalls and the AMPS argument motivating it are of a general nature, not tied to string theory, or any particular account of the nature or formation of the firewall. However, there is a string theoretical proposal along the lines of a firewall. Certainly this proposal falls into the category of "drama at the horizon", though it should be noted though that it predates the AMPS paper (e.g., Mathur [2005]).

Mathur [2009] shows that one cannot escape Hawking's argument by small quantum corrections adding a small amount of quantum 'hair' to the black hole, which might account for apparently lost information. Rather, avoiding information loss requires a great deal of quantum hair – a 'fuzzball' of such hair, in fact! The work of Mathur and collaborators explores a string theoretic model of just this kind, with significant consequences for spacetime in black hole models as we shall now explain – though not before noting that some

[15] See Susskind [2012a,b,c] for further discussions.

of this work is controversial, even in the string theory community, unlike the preceding.

We observed earlier that in modeling a stringy black hole one constructs a system, at a point of the large spatial dimensions, with the mass and charges necessary to produce it according to the supergravity EFE. Such modeling treats the strings and branes as sources for geometry, determining the background required to avoid the Weyl anomaly. It ignores any contribution that the strings themselves might make to the geometry, which after all is just a string state itself. Moreover, in the BPS (or adiabatic) calculations of entropy, the coupling is 'turned off', so the specific states counted are not gravitational at all. As a result, the construction is really only valid 'sufficiently far from' the strings, and does not tell us the geometry of the black hole in the vicinity of the strings – in particular about what happens close to the classical singularity.

Of course one would like to know that, but such ignorance does not immediately cast doubt on the rest of the construction. The string system is supposed to be at a point in the large spacetime dimensions in which the horizon forms, and prima facie, and it is reasonable to suppose that the size of the string system is no more than the Planck or string length, thus far from the horizon until the last stages of evaporation. Hence one expects that the derived geometry describes most of the black hole interior accurately; that string theory agrees with classical supergravity except near the singularity. However, explicit calculations of the string dynamics show that the stringy objects producing the black hole vibrate in its interior, so are not truly located at a point, but apparently extend to form a fuzzball. Thus the prima facie argument cannot be trusted, and one has to ask how large the vibrations are to discover how much of the black hole is occupied by the fuzzball.

Studying the fuzzball in detail will reveal the detailed string state, and hence the geometry of the region that contains it – if indeed the state of the fuzzball corresponds to a classical geometry at all. The calculation (reviewed in Mathur [2012]) exploits a duality between the stringy source of the black hole and a long, floppy string, to estimate the size of the vibrations, so of the fuzzball: around $(g^2\alpha'\sqrt{Q_1Q_5N/RV})^{-1/3}$ – which happens to be the radius of the horizon! Contrary to expectations, the fuzzball is not confined to the center, but apparently fills the black hole.

Applying the AdS-CFT correspondence, Mathur [2012] argues for the following picture (details are well beyond the scope of this paper). The fuzzball is not isotropic, as one implicitly assumes by taking it to be a point source in finding the supergravity black hole geometry; when anisotropy is taken into account the solution no longer has a horizon or singularity (Lunin et al. [2002]). In fact, the fuzzball causes the compact, cylindrical and toroidal, dimensions to 'cap off' at the horizon, though leaving the geometry away from the black hole essentially unchanged. Picture a cylinder smoothly tapering to a curved end; the open dimension ends where the circular dimension shrinks to a point, as there is no more space to travel into. Something similar happens around the fuzzball, although the geometry is more complicated (a 'Kaluza-Klein monopole'); the compact dimensions (S^1 and T^4 in our model) apparently cap off at the horizon radius, similarly terminating any trajectory into the black hole – a 'fauxrizon', marking the end of space external to the black hole!

What is beyond this fauxrizon? The results just quoted apply to individual states of the fuzzball; from that point of view there is no 'interior' strictly speaking, and no 'beyond' in a spatial sense, just a nonspatial, fundamentally stringy state. However, Mathur's group has shown how approximate spatial structure might be attributed to an 'interior', in terms of suitable statistical averages of states: internal space as a kind of thermodynamical property of the fuzzball. Thus if one asks, in a more operationalist spirit, what happen if you throw something through the fauxrizon, there are two possible answers. Perhaps the object 'sees' the thermodynamical space in the interior and passes through; in a more fundamental description, the result of some complex interaction with the fuzzball is that the object emerges on the other side, changed to reflect an apparent passage through it. Or perhaps, objects are simply amalgamated into the fuzzball state at the fauxrizon; after all, both fuzzball and matter are ultimately just complicated compositions of the same fundamental objects of string theory. In that case, operationally there truly is no interior – and there is plenty of 'drama' at the fauxrizon!

In either case, there is no horizon to cause an information paradox, and the fuzzball models recover both S_{BH}, and the Hawking radiation rate. But particularly in the latter case, we have an example where it is obviously inappropriate to ascribe a classical geometry to the interior, along the lines suggested by Polchinski earlier. Clearly in this case, Maudlin's construction does not apply; unitarity – and indeed information conservation – is obtained by the details of the fuzzball dynamics.

6 Conclusion: implications for the nature of spacetime and matter

In this paper we have reviewed how black holes can be modeled in string theory. While our main focus is on questions of ontology we also briefly addressed epistemological questions, arguing that the derivation of BHT from should be understood more as a consistency check, and weakly rather than as strongly confirmatory. We emphasize, however, that the importance of the models lies in giving a successful account of the underlying states, and so providing a Boltzmannian understanding of the entropy.

However, the main purpose of the paper is to throw some light on the ontological questions about spacetime and matter. To that end we have explicated the 'standard' interpretation of classical spacetime and matter according to string theory (§3): both classical matter and geometry correspond to coherent states of strings in suitable excitations. Then we described its application to a black hole model (§4), and investigated some of the possible implications for spacetime (§5): importantly, the fuzzball suggestion that an effective spacetime description breaks down at the 'horizon'. To conclude we will turn the focus back to the standard interpretation, and draw some lessons for the status of spacetime in string theory from our discussion.

First, the interpretation applies to the stringy black hole: Weyl symmetry leads to GR and classical supergravity, according to which the brane construction at a point produces a horizon in the spacetime geometry. Alternatively, if stringy matter is in fact a fuzzball, or there is a firewall, then the spacetime description breaks down at the 'horizon'; the geometry is as before outside, but the 'inside' is purely stringy. Either way, spacetime geometry is an effective description of a multi-string coherent state (and not a fundamental, classical geometry).

Second, empirical significance of the derived structure – the metric $g_{\mu\nu}$ – comes in the first place from its role in determining scattering amplitudes: it appears in the path integrals (2) and (3) and so different values lead to different cross-sections for observed particle scattering. However, stringy astrophysical models like black holes demonstrates further significance: astronomical observations of spacetime structure are understood as low-resolution observations of fundamental stringy fields. These points show that one has to be cautious with the claim that string theory has no empirical consequences: it reproduces the predictions of GR including observable objects like black holes (and scattering amplitudes, though not yet of the standard model). What it lacks (so far) are specific *novel* predictions, testable using current technologies.

Third, how cogent is the interpretation? The most questionable point concerns the existence (at least approximately) of suitable coherent states: string theory as developed is inherently perturbative, and the possibility of such states is postulated for an unknown exact theory. That is no argument against the picture, and indeed once the basic framework of perturbative string theory is accepted, it is a small step to coherent states; but the point does emphasize how the interpretation is speculative.

Fourth, supposing that coherent string states exist, and that they have an effective description as coherent states of quanta, one must ask about the classical limit: as a general question about QFT, do coherent states adequately explain the observed behaviors of classical fields? There is remarkably little discussion of this question in the literature[16], but one question in the present case is whether graviton coherent states remain coherent long enough to model cosmological scenarios? States will retain their coherence, and classical-like behaviour, only if their equations of motion are linear; so graviton coherent states will certainly lose their coherence, because of the non-linearity of the field equations. But on what time scales should we expect to see non-classical, quantum behaviour as a result? On the one hand, for a Schwarzschild black hole in the Wheeler-DeWitt framework Kiefer and Louko [1998] find the dispersion time to be $10^{73} \times (\text{mass in solar masses})^3$ seconds – a comforting 56 orders of magnitude greater than the age of the universe (and of the order of the Hawking radiation time) for a solar mass black hole! On the other, Wallace [2012, §3.3] points out that the chaotic nature of less symmetric gravitational systems can lead to a rapid loss of coherence. So matters are unclear.

Finally, we return to the suggestion made by Bokulich and Curiel regarding the relation between matter and physical geometry. According to the standard interpretation the 'conversion' of classical matter to geometry, and the reconversion of geometry back to matter in the form of quantum radiation is ultimately a transition between different multi-string states. In the first case from states of strings in a matter mode to states of strings in a graviton mode; in the latter, back from states of stringy gravitons to stringy matter

[16] Rosaler [2013] is a significant exception, and it is explored further in Huggett and Wüthrich [forthcoming].

quanta. Once again, with only perturbative string theory in hand one does not have a full theory of how these transitions occur. However, the mechanism (4) provides a model for what may occur; namely excitations of the branes inside the black hole decay into stringy quanta in the exterior. Such a process, and the modeling of black holes in string theory more generally, illustrates how the fungibility of geometry and matter is dynamical.

References

Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully. Black holes: complementarity or firewalls? *Journal of High Energy Physics*, 2013 (2):62, 2013.

Jacob D Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333, 1973.

Gordon Belot, John Earman, and Laura Ruetsche. The Hawking information loss paradox: The anatomy of controversy. *The British journal for the philosophy of science*, 50(2):189–229, 1999.

Karen Crowther, Niels S. Linnemann, and Christian Wüthrich. What we cannot learn from analogue experiments. *Synthese*, (1-26), 2019.

Erik Curiel and Peter Bokulich. Singularities and black holes. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2012 edition, 2012.

Radin Dardashti, Karim PY Thébault, and Eric Winsberg. Confirmation via analogue simulation: What dumb holes could tell us about gravity. *British Journal for the Philosophy of Science*, 68(1):55–89, 2017.

Sumit R Das and Samir D Mathur. The quantum physics of black holes: Results from string theory. *Annual Review of Nuclear and Particle Science*, 50(1):153–206, 2000.

Anthony Duncan. *The conceptual framework of quantum field theory*. Oxford University Press, 2012.

MB Green, JH Schwarz, and E Witten. *Superstring Theory*, volume I. Cambridge University Press Cambridge, 1987.

Daniel Harlow. Jerusalem lectures on black holes and quantum information. *Reviews of Modern Physics*, 88(1):015002, 2016.

Stephen W Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.

Stephen W Hawking. Breakdown of predictability in gravitational collapse. *Physical Review D*, 14(10):2460, 1976.

Gary T Horowitz, Juan M Maldacena, and Andrew Strominger. Nonextremal black hole microstates and u-duality. *Physics Letters B*, 383(2):151–159, 1996.

Nick Huggett. Target space $\neq$ space. *Studies in history and philosophy of science part B: Studies in history and philosophy of modern physics*, 2015.

Nick Huggett and Tiziana Vistarini. Deriving general relativity from string theory. *Philosophy of Science*, 82(5):1163–1174, 2015.

Nick Huggett and Christian Wüthrich. Emergent spacetime and empirical (in) coherence. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):276–285, 2013.

Nick Huggett and Christian Wüthrich. *Out Of Nowhere*. Oxford University Press, forthcoming.

K. Karaca. Kitcher's explanatory unification, kaluza-klein theories, and the normative aspect of higher dimensional unification in physics. *British Journal for the Philosophy of Science*, 63(2):287–312, 2012.

Claus Kiefer and Jorma Louko. Hamiltonian evolution and quantization for extremal black holes. *arXiv preprint gr-qc/9809005*, 1998.

Oleg Lunin, Juan Maldacena, and Liat Maoz. Gravity solutions for the d1-d5 system with angular momentum. *arXiv preprint hep-th/0212210*, 2002.

Samir D Mathur. The fuzzball proposal for black holes: An elementary review. *Fortschritte der Physik: Progress of Physics*, 53(7-8):793–827, 2005.

Samir D Mathur. The information paradox: a pedagogical introduction. *Classical and Quantum Gravity*, 26(22):224001, 2009.

Samir D Mathur. Black holes and beyond. *Annals of Physics*, 327(11):2760–2793, 2012.

Keizo Matsubara. Realism, underdetermination and string theory dualities. *Synthese*, 190(3):471–489, 2013.

Tim Maudlin. (information) paradox lost. *arXiv preprint arXiv:1705.03541*, 2017.

Thorne Misner and Kip S Thorne. *Wheeler, Gravitation*. Freeman, 1973.

Luboš Motl. What is background independence and how important is it?, 2012. URL http://motls.blogspot.com/2012/12/what-is-background-independence-and-how.html.

Joseph Polchinski. The black hole information problem. In *New Frontiers in Fields and Strings: TASI 2015 Proceedings of the 2015 Theoretical Advanced Study Institute in Elementary Particle Physics*, pages 353–397. World Scientific, 2017.

Joseph Gerard Polchinski. *String theory*. Cambridge university press, 1998.

James Read. On miracles and spacetime. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 65:103 – 111, 2019. ISSN 1355-2198. doi: https://doi.org/10.1016/j.shpsb.2018.10.002. URL http://www.sciencedirect.com/science/article/pii/S1355219818300704.

Joshua S Rosaler. *Inter-theory relations in physics: case studies from quantum mechanics and quantum field theory*. PhD thesis, University of Oxford, 2013. URL https://ora.ox.ac.uk/objects/uuid:1fc6c67d-8c8e-4e92-a9ee-41eeae80e145.

Kian Salimkhani. Quantum gravity: A dogma of unification? In Alexander Christian, David Hommen, Nina Retzlaff, and Gerhard Schurz, editors, *Philosophy of Science Between the Natural Sciences, the Social Sciences, and the Humanities*, volume 9 of *European Studies in Philosophy of Science*, pages 23–41. Springer International Publishing, 2018. ISBN 9783319725772.

Andrew Strominger and Cumrun Vafa. Microscopic origin of the bekenstein-hawking entropy. *Physics Letters B*, 379(1-4):99–104, 1996.

Leonard Susskind. The paradox of quantum black holes. *Nature Physics*, 2 (10):665, 2006.

Leonard Susskind. Complementarity and firewalls. Technical report, 2012a.

Leonard Susskind. Singularities, firewalls, and complementarity. *arXiv preprint arXiv:1208.3445*, 2012b.

Leonard Susskind. The transfer of entanglement: the case for firewalls. *arXiv preprint arXiv:1210.2098*, 2012c.

Leonard Susskind and James Lindesay. *An introduction to black holes, information and the string theory revolution*. World Scientific, 2005.

Leonard Susskind, Larus Thorlacius, and John Uglum. The stretched horizon and black hole complementarity. *Physical Review D*, 48(8):3743, 1993.

Washington Taylor. String field theory. In Daniele Oriti, editor, *Approaches to Quantum Gravity: Toward a New Understanding of Space, Time and Matter*, pages 210–28. Cambridge University Press, 2009.

Jeroen van Dongen, Sebastian De Haro, Manus Visser, and Jeremy Butterfield. Emergence and correspondence for string theory black holes. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 69:112–127, 2020.

Tiziana Vistarini. *The Emergence of Spacetime in String Theory*. Routledge, 2019.

Spenta R Wadia. A microscopic theory of black holes in string theory: Thermodynamics and hawking radiation. *CURRENT SCIENCE-BANGALORE-*, 81(12):1591–1597, 2001.

Robert M Wald. *Quantum field theory in curved spacetime and black hole thermodynamics*. University of Chicago press, 1994.

David Wallace. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press, Oxford, 2012.

David Wallace. The case for black hole thermodynamics part i: Phenomenological thermodynamics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 64:52–67, 2018.

David Wallace. The case for black hole thermodynamics part ii: statistical mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 66:103–117, 2019.

David Wallace. Why black hole information loss is paradoxical. In Nick Huggett, Keizo Matsubara, and Christian Wüthrich, editors, *Beyond Spacetime: The Foundations of Quantum Gravity*, chapter 10, pages 209–36. Cambridge University Press, 2020.

Edward Witten. Reflections on the fate of spacetime. *Physics Today*, pages 24–30, April 1996.

Christian Wüthrich. Are black holes about information?, 2017.

Barton Zwiebach. *A first course in string theory*. Cambridge university press, 2004.