From time-irreversibility in gravity to measurement in quantum mechanics

Warren D. Smith
WDSmith@fastmail.fm

3 October 2003

Abstract — **1. We prove a theorem that says, essentially, that it is impossible for an infinite slab to exist which absorbs, reflects, or transforms a fraction>0 of the energy in weak gravitational waves. But it is (essentially) possible to make a 100% efficient emitter of such waves.**
2. We prove a theorem that says, essentially, that General Relativity (GR) is soluble forward in time but not backward in time, in a manner like the 1D heat equation $F_t = F_{xx}$.
3. We discuss other (known) time-direction asymmetries of GR. All of these ultimately arise, essentially, from the positivity of mass.
Hence we argue that, contrary to common false claims, gravitational physics is *not* time-reversal invariant and defines a unique direction of time as an *output* not input. Making this argument involves resolving several subtle apparent paradoxes. But non-gravitational physics (quantum field theories) obey charge-parity-time reversal invariance CPT. We conclude with the grand Manifesto that all the time-irreversibility phenomena observed in everyday life, including "quantum measurement," the 2nd law of thermodynamics, and outgoing-only radiative boundary conditions, ultimately arise from gravitational physics. We outline how that happens and discuss the implications about future quantum gravity theories.

Keywords — Topology, gravitational waves, irreversibility, partial differential equations, quantum mechanics, decoherence, measurement, arrow of time.

PACS #s: 04.20.-q, 04.30.-w, 04.60.-m, 03.65.Yz.

1 Absorbers or reflectors of gravitational waves

L.Smolin [42][43] presented arguments for the impossibility of a high-efficiency absorber or reflector ("shield") for gravitational waves. His arguments basically are a collection of designs for physical devices intended to be gravitational shields, together with analysis of each indicating they won't work. (Either because: their absorption or reflection coefficients are too small, or because they will collapse into a black hole, or because at best they'll be unstable to such a collapse.) But (1) a collection of failed designs does not prove that *every* possible design must fail, and (2) Smolin's arguments that some of his designs fail, seem incompletely convincing and at least one of his claims has been refuted [14]. So while I do *not* agree that Smolin "demonstrated impossibility," I *do* suspect that he is right – that there is some rigorous sense in which good gravitational wave reflectors and absorbers are forbidden. But what is it?

We will attack this problem in a completely different manner – geometrical/topological rather than physical.

For background about gravitational waves, including the linearization of Einstein's field equations and the facts that the length-oscillations of gravity waves in vacuum are always *traceless* and always *transverse* to their direction of propagation ("TT gauge") see pages 946-955 of [30].

As a warm up, consider a gravitational wave impinging on the top surface of a magic slab. The wave, due to some combination of absorption and reflection, does not affect the bottom surface of the slab, which therefore matches the metric of flat space exactly. To make matters concrete, set up a t, x, y, z coordinate system and let the top of the slab be $z = \ell$ and the bottom be $z = 0$. Let the metric inside the slab $(0 < z < \ell)$ be

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + (1 + \epsilon F(z)\sin t)\mathrm{d}x^2 + \frac{\mathrm{d}y^2}{1 + \epsilon F(z)\sin t} + \mathrm{d}z^2 \quad (1)$$

where $F(z)$ increases from $F(0) = 0$ to $F(\ell) = 1$ and where the gravitational wave amplitude ϵ is regarded as small. Such a metric would get the job done – at $z = 1$ there are volume-preserving oscillations of transverse distances (in the x and y directions) proportional to $\epsilon \sin t$, whereas at $z = 0$ we just have flat space. This metric has Einstein tensor ${}^{t}_{t}$ component given by the following exact formula:

$$G^t_t = \frac{\epsilon^2}{4} \frac{F(z)^2 \cos^2 t + F'(z)^2 \sin^2 t}{(1 + \epsilon F(z)\sin t)^2}. \quad (2)$$

Note this is always non-negative. Under Einstein's field equations of general relativity that corresponds to an *always non-positive mass-energy density.* Presumably such matter is forbidden – or at least, isolated chunks of it are forbidden – because if they could exist, presumably the vacuum would decay into such chunks, which experimentally does not happen. Under that assumption (that isolated negative density matter is forbidden) we conclude that a slablike gravitational shield of this (very specific) sort is impossible. The same argument may be reinterpreted as the assertion that a slablike *unidirectional emitter* of gravity waves (a.k.a. a "gravitational wave rocket") – again of this very specific sort – is impossible.

The argument so far has not been very impressive because it only applies to the very specific, very homogeneous metric of EQ 1. What about a gravity-wave absorber made of *in*homogeneous material? We shall now make a similar, but much more general, argument.

Theorem 1. *ANY twice-differentiable and band limited spacetime metric tensor which is within $\pm O(\epsilon)$ of the metric*

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

of flat spacetime for $0 < z < \ell$, and which matches flat spacetime on $z = 0$, and matches

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + [1 + \epsilon \sin t]\mathrm{d}x^2 + \frac{\mathrm{d}y^2}{1 + \epsilon \sin t} + \mathrm{d}z^2$$

on $z = \ell$ (both matches to within $O(\epsilon^2)$) must (for all sufficiently small $\epsilon > 0$) have negative mass.

Proof: We are first going to prove this under the assumption that the metric tensor $g_{\alpha\beta}$ is such that either $g_{xx} + g_{yy}$, or all three of g_{xx}, g_{yy} and g_{zz}, are monotonic functions of z when t, x, y-averaged. Then we will prove it without that assumption.

To accomplish the first goal we shall need only three ideas:

1. The equations $G_\alpha^\beta = \chi T_\alpha^\beta$ of general relativity[1] in the limit $\epsilon \to 0$, become the *linear* equations
$$-\overline{h}_{\alpha;\mu}^{\beta;\mu} = 2\chi T_\alpha^\beta \qquad (3)$$
as is explained on pages 435-438 of [30]. Here $h_{\alpha\beta}$ is the perturbation of the metric tensor away from that of flat space, the overline denotes a certain linear operation, and semicolons denote covariant derivatives (which in the present approximation are the same as ordinary derivatives). Therefore linear superpositions of metric perturbations h_α^β cause linear superpositions of the energy-momentum tensor T_α^β.

2. By taking positive linear combinations of an arbitrary band-limited[2] metric of bounded norm in our slab with various translations of itself in the x, y and t directions, we can make it assume a very homogeneous form in which the departures from the flat-space metric tensor are proportional to $\epsilon F_{\alpha\beta}(z) \sin t + \epsilon H_{\alpha\beta}(z) \cos t$ for some symmetric matrices F and H depending on z. The point is that there now is no dependence on x or y, and the only dependencies on t are proportional to $\sin t$ and $\cos t$, and not to, e.g. $\sin(3t)$.

3. By gauge invariance[3] we may demand that the t row and t column of F and H are all 0s, i.e. $F_{0\beta} = F_{\alpha 0} = H_{0\beta} = H_{\alpha 0} = 0$. (It is not strictly necessary to make this demand, but it is convenient because it greatly simplifies the symbolic manipulations that will follow.)

We now exactly compute the Einstein tensor $\begin{smallmatrix} t \\ t \end{smallmatrix}$ component G_t^t and then t-average it over $0 \le t < 2\pi$. The result (in the case $H_{\alpha\beta} = 0$) is

$$\begin{aligned}\frac{8}{\epsilon^2}\langle G_t^t \rangle = &(F_{13}^2 + F_{12}^2 + F_{23}^2) - (F_{22} + F_{11})F_{33} - F_{11}F_{22} \\ &-2[F_{22}''F_{33} + F_{22}''F_{22} + F_{11}''F_{33} + F_{11}''F_{11} + 2F_{12}''F_{12}] \\ &-[F_{22}'F_{33}' + F_{11}'F_{33}' + F_{22}'^2 + 3F_{12}'^2 + F_{11}'^2 - F_{11}'F_{22}'] \\ &+O(\epsilon^2)\end{aligned} \qquad (4)$$

where the primes denote z-derivatives, the angle brackets denote time-averaging, and the indices of t, x, y, z are $0, 1, 2, 3$ respectively. (To compare this to the previous EQ 2, use $F_{11} = -F_{22}$ in place of that EQ's F and make all our other $F_{\alpha\beta}$ be 0.)

As boundary conditions, we demand that $F_{33} = F_{12} = F_{13} = F_{23} = 0$ at $z = 0, \ell$ and $F_{11} = F_{22} = 0$ at $z = 0$. These demands, and $F_{11} = 1$ and $F_{22} = -1$ at $z = \ell$, are necessary just to make the metric tensor continuous.

The terms in the first line of EQ 4, i.e. those depending on the $F_{\alpha\beta}$ directly, i.e. not on their derivatives, are all 0 at $z = 0$ and are obviously positive near $z = \ell$ since all of them are 0 at $z = \ell$ except for $-F_{11}F_{22} = 1$. And of course all the three square terms are always nonnegative. Now suppose that the diagonal metric perturbations $F_{11}(z)$, $F_{22}(z)$ and $F_{33}(z)$ are *monotonic* functions of z, or that $F_{11}(z) + F_{22}(z)$ is monotonic. Then *all* the terms in the first line of EQ 4 are nonnegative at *every* z with $0 \le z \le \ell$, because $F_{33} \equiv 0$ is forced by its zeroness at the boundaries and by monotonicity (or because $F_{11} + F_{22} \equiv 0$), because squares always are nonnegative, and because $F_{11}(z)F_{22}(z) < 0$ by monotonicity and the boundary conditions. So these terms correspond under the Einstein field equations to a forbidden everywhere non-positive mass density; so far, so good.

Now consider the remaining terms on the right hand side of EQ 4. Integrate across the slab from $z = 0$ to

[1] $G_\alpha^\beta = R_\alpha^\beta - \frac{1}{2}R\delta_\alpha^\beta$ is Einstein's tensor here expressed in terms of the Ricci curvature tensor R_α^β, the curvature scalar $R = R_\mu^\mu$, and the Kronecker delta symbol; T_α^β is the energy-momentum tensor of the matter; and χ is 8π times Newton's gravitational constant, in units where lightspeed $c = 1$.

[2] Any 2π-periodic function of time is expandible in a Fourier series and saying that function is "band limited" means the Fourier series terminates. We then only need a *finite* "linear combination" of various t-shifts to make all the higher-frequency terms cancel out, and due to this finiteness, ϵ will at most be increased by some constant factor (dependent on the band cutoff). Meanwhile, the linear combination over x and y shifts can simply be averaging.

[3] Using a different gauge (which 't Hooft [18] calls the "temporal gauge") than the TT gauge used to present EQ 3; that gauge and the precise form of those equations will not be used further herein.

$z = \ell$, using integration by parts. We get

$$\int_0^1 [F'_{11}F'_{22} + F'_{22}F'_{33} + F'_{11}F'_{33} + F'^2_{22} + F'^2_{12} + F'^2_{11}]\mathrm{d}z \quad (5)$$

In the integration by parts, we dealt with the $F''_{11}F_{11} + F''_{22}F_{22}$ terms via the integration by parts formula $\int_0^\ell F''F\mathrm{d}z = (F^2/2)'|_0^\ell - \int_0^\ell (F')^2\mathrm{d}z$ and we impose the further boundary condition demand

$$\{F_{11}(z)^2 + F_{22}(z)^2\}']_0^\ell \le 0. \quad (6)$$

In this demand "= 0" would be expected if there is no power-dissipation in the gravity wave right at the slab-boundaries, which is what we would expect if our wave-absorbing device were located in the *strict* interior of the slab, with the rest being vacuum. If "<" holds (as would be expected if there is more wave-absorption at $z = \ell$ than at $z = 0$, which would be expected since the amplitude is supposed to be larger at $z = \ell$) then EQ 5 will only be a lower bound on the true integral, which will only make our argument more true.

In EQ 5, all the squares in the integrand are non-negative. Again, if $F_{11}(z)$, $F_{22}(z)$ and $F_{33}(z)$, or $F_{11}(z)+F_{22}(z)$, are monotonic, then $F'_{11} \ge 0$, $F'_{22} \le 0$ and $(F'_{11} + F'_{22})F'_{33} = 0$ are forced so that the other terms in the integrand also are all non-negative everywhere.

Essentially because sine and cosine behave the same, the calculation would be essentially the *same* if we had set $F_{\alpha\beta} = 0$ instead of $H_{\alpha\beta} = 0$, but a little easier, since the boundary conditions $H_{\alpha\beta}(z) = 0$ at $z = 0, \ell$ for all α, β are simpler. We shall not give the cloned argument explicitly, and instead shall merely note that the same result (positivity of $\int_0^1 8\langle G^t_t\rangle\mathrm{d}z$) follows from it.

Finally, essentially because of the orthogonality of $\sin t$ and $\cos t$ and the properties of integrals of quadratic functions of $\sin t$ and $\cos t$, we may combine our twin results about $\int_0^1 8\langle G^t_t\rangle\mathrm{d}z$ to conclude that it must be positive for *arbitrary* $F_{\alpha\beta}$ and $H_{\alpha\beta}$ (*both* of which are now allowed to be nonzero), provided the 3 diagonal terms g_{11}, g_{22}, and g_{33}, or just $g_{11} + g_{22}$, are required to be monotonic functions of z (after t, x, y-averaging within the slab).

Due to the *strict* positivity, all this also is true if we relax the z-monotonicity demands to merely that the relevant functions be sufficiently *close* to being monotonic.

To relax the monotonicity requirements *completely*, we need two new ideas. One is to consider the transverse trace of the perturbation to $g_{11} + g_{22}$, which note is zero when $z > \ell$ and $z < 0$ (at least, to accuracy $O(\epsilon^2)$). We want to argue that in fact, it must be zero (and hence monotonic) for all z, and hence that $F_{11} + F_{22} = H_{11} + H_{22} \equiv 0$, so that the theorem applies.

To do this, recognize that $F_{11} + F_{22}$, as a function of z, is proportional to the cross sectional area of space at that z, which is oscillating proportional to $\sin t$.

If our infinite slab absorbs a nonzero fraction of the incoming gravity wave, then we can slice it into finite square tiles , i.e. instead of $-\infty < x < \infty$ and $-\infty < y < \infty$, we restrict x and y to large finite intervals. If these tiles are made large enough, then cutting them apart will be unable to significantly reduce the efficiency of the absorber (just by considering the small size of the region near the cut, and considering speed of light limits). Once these tiles are cut free, we can rotate them by any multiple of 90°, and/or mirror-reverse them, and the chunks (as wave absorbing mechanisms) should still work essentially as well (since our gravity wave is symmetric[4] under interchange of x and y coordinates). So, if we slice our slab into an infinite number of square tiles and rotate each tile 90° – or not – and mirror reverse it – or not – depending on independent coin flips, then, with probability 1, we will cause $F_{11} + F_{22} \to 0$. That is merely a matter of phase cancellation: If $F_{11} + F_{22} > 0$ at some z, that means that space has increased cross sectional area when $\sin t > 0$, and recall that the incoming gravity wave has increased x-lengths and decreased y-lengths when $\sin t > 0$. However, our gravity-wave absorbing slab, due to the tile-randomization trick, *no longer knows* which coordinate is x and which is y. So it does not "know" whether to have increased cross sectional area when $\sin t > 0$ or when $\sin t < 0$; hence must average it out to zero.

Therefore, $F_{11} + F_{22}$ becomes arbitrarily close to a monotonic function, namely 0, and the theorem now holds. Specifically, let "mass" mean the density (which in general relativity with $- + ++$ signature is proportional to $-G^t_t$) averaged both over time t and spatially over z across the width of the slab $0 < z < \ell$. We conclude that *any* slab-shaped absorber, reflector, and/or unidirectional emitter of band-limited gravity waves, in the limit $\epsilon \to 0$ in which space is nearly flat, must have negative mass. Q.E.D.

Our proof above indeed goes beyond merely showing the impossibility of a non-negative mass slab with transverse length oscillations on its top side but none on its bottom.

Extension: *It shows the impossibility of a nonnegative-mass slab with* different *transverse length oscillations on its top or bottom.*

This argues for the impossibility of extracting, reflecting, or transforming (the amplitude or polarization of), *any* nonzero constant fraction of the power in impinging plane-parallel sine waves.

Due to limitations of our proof technique, we were only able to establish this[5] in the linearized gravity weak-field

[4] To accuracy $O(\epsilon^2)$, which suffices for us.

[5] This nevertheless seems adequate for practical purposes since almost all of our universe *is* nearly flat. E.g., even inside the sun $\epsilon < 0.0001$. Thus, only tiny amounts of weak plane gravity waves (corresponding to the tiny amounts of our universe that are highly curved) at most can be absorbed or reflected in the present universe, and this will remain true at least up until the moment (which probably will never come) when our universe becomes highly curved.

If there is a nonzero cosmical constant Λ then that *can* give the effect of negative mass and consequent red shifting *can* cause "absorption" of gravitational (and other) waves (although the geometry is now nonEuclidean and the notion of a "slab" loses meaning). However, that kind of absorption still does not allow metrical riplimit where the metric was demanded to be within ϵ of the metric of flat space as $\epsilon \to 0$, and for plane waves impinging on a slab.

But we believe that the fundamental impossibility here is of a *topological* nature. To make a crude analogy, it is impossible to have a 2D Riemannian metric which smoothly joins the surface metric of a cone to that of another cone pointed the other way, at a common circle of radius 1, unless, somewhere, there is negative curvature. That is an obvious topological fact. (It may be proven with the aid of the 2D Gauss-Bonnet theorem.) Our theorem seems to be a much less obvious, but somehow similar, fact about joining two gravitational plane-wave metrics. We therefore **conjecture** that this result (perhaps if weakened to assert negative mass of only some *part* of the slab) is valid much more generally, without need for ϵ to be small. But formulating and proving such a conjecture is difficult, e.g., it is hard even to define "mass" in a highly non-flat metric.

1.1 Impossibility of extracting energy from gravity waves – paradox?

But doesn't all this contradict the common wisdom that it is possible to build gravitational wave detectors with nonzero efficiency? Actually, it doesn't, although understanding how that can be, is tricky. We proceed in 4 steps to reach the point where we can see the reason paradox is evaded.

(1) All the most popular kinds of detectors built by the experimentalists in a (so far unsuccessful) effort to detect gravity waves do not work (as far as I can tell) by extracting energy per se from a gravity wave.

(2) "Plane-parallel wave" metrics

$$\mathrm{d}s^2 = H(u,x,y)\mathrm{d}u^2 + 2\mathrm{d}u\mathrm{d}w + \mathrm{d}x^2 + \mathrm{d}y^2 \qquad (7)$$

are exact solutions of the Einstein vacuum field equations if $H_{xx} + H_{yy} = 0$ (these subscripts denote partial differentiation). As may be seen by computing the Riemann curvature tensor $R^{\alpha}{}_{\beta\mu\nu}$, these are non-flat if H_{xy}, H_{xx}, or H_{yy} are nonzero. However, all their polynomial curvature invariants, such as R, $R_{\alpha\beta}R^{\alpha\beta}$, $R^{\beta}_{\alpha}R^{\mu}_{\beta}R^{\alpha}_{\mu}$, and $C_{\alpha\beta\mu\nu}C^{\alpha\beta\mu\nu}$ are zero, the same as flat space. Furthermore, all their first-order differential curvature invariants, such as $R_{\alpha\beta;\mu}R^{\alpha\beta;\mu}$ and $C_{\alpha\beta\mu\nu;\kappa}C^{\alpha\beta\mu\nu;\kappa}$, also are zero. So any sort of gravitational wave detector that reacts only to these invariants, will be unable to distinguish these gravitational waves from flat space.

(3) But nevertheless, in principle, it *is* possible to extract energy from a gravity wave. Position two masses M_1 and M_2 near opposite ends of a highly rigid meter stick. Suppose a gravity wave alters the distance between M_1 and M_2 to, say, 99cm ($\epsilon = 0.01$). Meanwhile, the two ends of the meter stick will remain 100cm apart – if the frequency f of the impinging wave is much smaller than the elastic resonant frequencies of the stick. Therefore, if a small amount of sliding friction is supposed between the masses and the meter stick, we will generate heat. The force experienced by M_1 and M_2 is proportional to their masses and hence the heat-power extractible is proportional to $M\epsilon\ell f$ where ℓ is the length of the stick. Note that ℓ cannot be arbitrarily large since $\ell f < c$ is required even for a stick made of the stiffest imaginable stuff, whose speed of sound is the same as lightspeed c. Therefore, with the parameters ϵ and f of the wave being fixed, to extract nonzero power it is necessary that M be bounded above zero.

(4) Now suppose we make an infinite plane sheet covered with the energy-extractor devices. This is intended to intercept some nonzero constant fraction of the energy in a gravity wave of some constant frequency impinging normally on our sheet. I.e. this is intended to extract some nonzero amount of power per unit area. Accomplishing that requires the areal mass density of our covered sheet to be bounded above zero. *But* then the amount of mass in a large disk of radius r drawn on our sheet grows proportionally to πr^2, whereas the maximum mass that is allowed to exist inside a ball of radius r is proportional to r (since the Schwarzschild radius of a mass M is proportional to M, not $M^{1/3}$ as might naively have been expected)! For all sufficiently large r, that is a contradiction! So: no spacetime metric can exist of this type, and certainly not one that satisfies our assumptions that it is within ϵ of flatness.

Conclusion: So while devices that extract energy from a gravitational wave are possible, that does *not* contradict the claimed impossibility of extracting any constant fraction of the power in an infinitely wide plane sine wave, by means of a (possibly infinite) nonnegative-mass device contained in an infinite slab $-\infty < x < \infty$, $-\infty < y < \infty$, $0 < z < \ell$ in approximately-flat space.

2 Introduction to Unidirectionality of time & thermodynamics

Smolin [42][43] had drawn some far-reaching conclusions about entropy from his arguments against gravity wave absorbers. We are going to draw related but even further-reaching conclusions.

Observe that matter can lose information and energy by radiating gravitational waves. In fact that can be quite efficient: two orbiting black holes will eventually spiral into each other due to loss of energy via gravitational wave emission. If one of the holes has much smaller rest mass than the other and is traveling arbitrarily close to lightspeed on an initial trajectory arbitrarily near to the "photon orbit," then in fact this process will radiate away a fraction arbitrarily near 100% of the initial energy and angular momentum. That radiation, viewed at very large distances, would look like a weak plane wave.

But, due to our impossibility proof for good absorbers for weak plane gravity waves, only a *very* tiny fraction

ple information to "come back" to the domain of ordinary matter (since Λ is not ordinary matter, and the information is still all there, merely red shifted) and hence presumably the possibility that $\Lambda \neq 0$ does not affect the arguments we shall make in the remainder of this paper.

of that lost information and energy will ever be able to "come back" from the world of metrical waves to the world of matter.

There is asymmetry here. Black holes can spiral into one another and merge, but not bifurcate and spiral out from one another. Good emitters of weak plane gravity waves can exist, but good absorbers cannot. Information can be lost by matter into metrical ripples, but not regained.

We now see that the latter two sentences are consequences purely of (1) Einstein's field equations and (2) the positivity of mass. Thus it could be argued that the unidirectionality of time and the 2nd law of thermodynamics (regarded as a statement of the form "information can be lost, but not gained") are consequences purely of these two things.

Note also that this is a fundamental difference between gravitational and nongravitational physics. E.g. very good reflectors, polarization-rotators, and so forth clearly can exist for bandwidth-limited electromagnetic plane waves, if gravitational effects are neglectible – and these devices experience no difficulty having positive mass.

However, if we now do *not* neglect gravitational effects, and if we assume that any electromagnetic plane wave must automatically be coupled to a co-propagating gravitational plane wave (exact solutions [24] of this sort for the Einstein-Maxwell equations are known) it then follows that it is impossible to build an infinite-slab-shaped reflector or absorber for weak electromagnetic plane waves!

3 Irreversibility of the general relativity PDEs

Everybody who has taken a course in Partial Differential Equations [20] has encountered the fact that the 1D "heat equation" $F_t = F_{xx}$ is uniquely soluble, starting from any initial twice-differentiable $F(x)$ whatever, going forward in time for any timespan whatever. $F(x)$ any nonzero time later is infinitely differentiable and depends continuously on the initial $F(x)$. *But* this same equation *cannot* be solved going backward in time – no matter how small a timespan we choose, in general no solution will exist! Heat equation time evolution is in some sense a "many parents to one child" map, and with many "orphan children," and thus is inherently irreversible.

The reason this is so is as follows. (For simplicity, we shall only demonstrate it for 2π-periodic even functions of x.) Let a_0, a_1, a_2, etc. each be i.i.d. real random variates selected uniformly from any fixed finite-size open interval. Then the 2π-periodic even k-tuply-differentiable function

$$F(x) = a_0 + \sum_{n\geq 1} \frac{a_n}{n^{k+2}} \cos(nx) \tag{8}$$

is soluble going forward in time by t:

$$F(x,t) = a_0 + \sum_{n\geq 0} \frac{a_n}{n^{k+2}} \cos(nx) \exp(-n^2 t) \tag{9}$$

and observe that this series always converges for all $t \geq 0$. As $t \to +\infty$ notice that $F(x)$ becomes flat: $F(x,t) = a_0 + O(e^{-t})$. But observe that this series (with probability 1) *diverges* at every $x \neq 0$ for *any* $t < 0$. Nor is there any analytic continuation to any $t < 0$; the whole imaginary t-axis is a so-called "natural boundary" to analyticity. If we try to go backwards to negative time, then $F(x)$ "instantly roughens infinitely much" preventing the 2nd spatial derivative F_{xx} from even existing anywhere, and hence preventing the PDE from even being applicable.[6]

Oddly, I have never before seen stated the fact that the PDEs of *general relativity* obey exactly the same sort of time-irreversibility and many-to-one properties.

First of all, GR has a unique solution going forward in time, in *some* future-open set (perhaps not extending very far into the future, but a positive amount everywhere) containing the initial 3-metric, starting from sufficiently nice initial conditions. ("Sufficiently nice" means satisfying certain equality constraints and Sobolev-type norm bounds.) That was proven by Y.Choquet-Bruhat [9][46]. The future-time solution depends continuously on the initial data. Furthermore, from sufficiently nice initial conditions sufficiently close (in certain Sobolev-norm senses) to the metric of flat $(3+1)$D spacetime, the equations of *vacuum* GR have a unique solution forward in time *eternally*, depending continuously on the initial data, and this solution ultimately tends (in certain norm senses) to the metric of flat space as $t \to +\infty$. That was proven by Christodolou and Klainerman [10] in a prizewinning 514 page proof.

Klainerman and Nicolo [22], after redoing the Christodolou-Klainerman proof by different methods in about half the number of pages, further showed from sufficiently nice initial conditions sufficiently close (in certain Sobolev-norm senses) to a Schwarzschild black hole metric

$$ds^2 = \frac{2M-r}{r} dt^2 + \frac{r}{r-2M} dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{10}$$

with some constant mass $M \geq 0$, the equations of vacuum GR have a unique solution forward in time eternally, which tends as $t \to +\infty$ to a Schwarzschild metric.

The *linearized* perturbation analysis of the flat space and Schwarzschild metrics had been done well before Klainerman et al. did the fully nonlinear analysis, and had, of course, indicated the same results. In particular, recall our Fourier analysis of the 1D heat equation above, and the key fact that each Fourier mode decays exponentially with time, with the slowest decay rate being $\exp(-t)$, achieved by the $n = 1$ mode. Similarly the Schwarzschild metric perturbations are (under the appropriate linearization of the GR field equations) expressible as linear combinations of "quasimodes," each

[6] Although as written, this example concerns C^k initial data, one may easily extend it to concern C^∞ initial data by replacing the convergence factor n^{k+2} in EQ 8 by any super-polynomially, but sub-exponentially growing function of n, such as $\exp\sqrt{n}$.

of which decays oscillatory-exponentially with time, i.e. like $\text{re}\exp(-Kt)$ where K is a complex number with $\text{re}K > 0$. This is discussed in [1] [3] [4] [6] [7] [8] [23] [25] [26] [27] [28] [29] [31] [33] [34] [39]. The quasimodes are indexed by two integers n and ℓ in the same manner as the spherical harmonics, so there are a countably infinite number of them, indeed for each ℓ there are an infinite number [4]. The characteristic decay times and the oscillation periods both are equal to $c^{-3}G_{\text{Newton}}M$ (where M is the hole's mass) times a function (which we may regard as assuming *complex* values) of n and ℓ. Call this function $1/f_{n\ell}$. It was recently proven [31][32][2] (confirming previous numerical observations [1][33]) that as $n \to \infty$,

$$4f_{n\ell} = 2\pi i(n - 1/2) + \ln 3 + O(n^{-1/2}) + iO(\ell), \quad (11)$$

which note indicates an asymptotically *constant* ringing frequency $\text{re}f$, even as the angular-spatial mode number n is taken to ∞ and the decay rate (proportional to imaginary part of f) also goes to infinity. There is a positive minimum decay rate $\min_{\ell n} \text{im} f_{\ell n} > 0$ among these. The fundamental tone of a Schwarzschild hole of 10 solar masses, according to [23], corresponds to a frequency of 1.2kHz and a damping time of 0.55mSec.

It therefore immediately follows (in a manner exactly analogous to our Fourier analysis of the 1D heat equation) that Schwarzschild-perturbations are generally *not* soluble backward in time for *any* nonzero amount of time, no matter how small. Suitable randomized infinite sums of quasimodes which are k-tuply differentiable (for any desired $k > 0$) will, in fact, be soluble forwards in time eternally even under full nonlinear GR, but (with probability 1) not soluble backwards in time (under vacuum GR) for any nonzero timespan whatever. We summarize this as a theorem.

Theorem 2. *The Einstein PDEs of vacuum General Relativity have a unique solution, depending continuously on the initial data, forwards in time for some nonzero amount of time starting from each initial point. (The amount of time may depend on the point. "Sufficiently nice" means satisfying certain equality constraints and Sobolev-type norm bounds.) If the initial data is a sufficiently small-norm perturbation of a Schwarzschild black hole metric, then this unique solution exists eternally. But such metrics (even if they are demanded to be C^k for any fixed integer k, including $k = \infty$ as in footnote 6) generically have no solution that anywhere goes backwards in time for any nonzero timespan whatever.*

4 Other time-irreversibility properties of General Relativity

Well known theorems of Hawking and Penrose [16][5] show that, in General Relativity in which the matter (if any) obeys suitable "energy conditions" (and the cosmical constant is zero)

1. The surface area of the horizons of black holes can increase, but never decrease.
2. Black holes can merge but never bifurcate.
3. Any $(3 + 1)$-metric having a "trapped surface" will, under General Relativistic time evolution, inevitably develop a singularity sometime in the future.

The singularity theorem has recently been reproven [36] without requiring the "strong" or "dominant" energy condition [16], but instead merely requiring the "averaged null energy condition." All of these energy conditions are, essentially, various expressions of the notion that "mass is positive." That causes mass-energy to move at sublight speeds and causes nearby timelike geodesics to converge toward one another rather than diverge apart (in the proofs by Hawking, this is the fundamental time-asymmetry property on which everything else rests). All three theorems remain interesting even in *vacuum*, in which case they hold entirely independently of any assumptions, including energy conditions, about matter. In that case, the fundamental time-asymmetry property on which everything else rests is instead the topological-geometrical properties of "trapped surfaces." The fact that trapped surfaces are oriented future-inward may also be regarded as a statement about positivity of mass.

It should now be clear that the unidirectionality of time is intimately related to the positivity of mass. Therefore, it is worth noting the "positive mass theorem," which is a purely topological claim about asymptotically flat $(3 + 1)$-metrics [48]. Namely, it states that asymptotically flat $(3 + 1)$-metrics without metrical singularities have nonnegative ADM mass and nonnegative Bondi mass provided[7] the Einstein tensor everywhere obeys certain energy conditions.

5 The apparent time-reversibility of General Relativity – paradox?

The above theorems show that general relativity has built in time-*ir*reversibility properties.

But wait! The whole of GR is based on the metric tensor $g_{\alpha\beta}$ and its associated quadratic form ds^2, the squared infinitesimal length element. If we reverse the sign of all 4 of the coordinates t, x, y, z, then all possible quadratic terms, e.g. dt^2, dx^2, $dx dt$, and $dx dy$, remain the same. Hence ds^2 and the metric tensor remain the same. Hence every curvature tensor remains the same, although Christoffel symbols and covariant derivatives reverse sign. Thus, both the Einstein field equations, and the geodesics (regarded as point sets), are invariant under "P^{3}T" (where T is time-reversal and P is a spatial mirror reversal) provided the matter tensor T_α^β is left unchanged (or if we are in vacuum, because $T_\alpha^\beta = 0$ it need not concern us).

[7] ADM and Bondi mass are based on the far-distance behavior of the gravitational fields. An extension of this theorem allows metrical singularities to exist provided they are hidden behind event horizons.

Therefore, any solution of GR is also a solution after P^3T. Is this a paradox?

The resolution of this paradox is again subtle. It is again necessary to increase our understanding in stages before seeing in stages 3 and 4 the whole picture of why this is not really a paradox.

(1) The "new" solution is really just exactly the same metric as the "old" solution – we have merely made a coordinate change – and a coordinate change whose Jacobian is the negated 4×4 identity matrix, which is *connected* to the identity matrix by a continuous path of 4×4 orthogonal matrices with $+1$ determinant – which in the view of GR is no change at all.

If you say you have a solution of GR which involves a black hole bifurcating into two, I will reply that your solution really describes two black holes merging into one and I will supply a continuous path of infinitesimal coordinate changes to your notions of space and time which will cause those notions to turn into my notions. Then I will insist that my notions of space and time are the correct ones. (Of course, you might now complain that this "insistence" of mine was arbitrary – why don't we instead insist on your notion of time? But in fact, as we shall see at stage 3, it isn't arbitrary.)

Saying that GR obeys a P^3T symmetry is a tautology since this operation has no effect on a gauge-canonized metric. If the $(3 + 1)$-manifold we live in is oriented (as it must be if time is unidirectional) then performing P^3T is to do nothing.

(2) A critic might argue that ours are not *really* time-irreversibility theorems because, really, the time-irreversibility was not proven but rather *assumed* in the form either of energy conditions or trapped surface conditions. For example (the critic might continue), the time-reverse of the "black holes can merge but not bifurcate" theorem would be "white holes can bifurcate but not merge." The time reverse of the theorem that small perturbations of a Schwarzschild black hole metric die out exponentially, would be that small perturbations of a Schwarzschild white hole (i.e. negative mass) metric increase exponentially. If we had instead assumed the *negativity* of mass, then nearby timelike geodesics would tend to diverge. The time reverse of a ball of gas held together by gravity but prevented from collapsing by internal pressure, would be a ball of negative-mass particles held together by negative pressure (i.e. tensile stress) but prevented from imploding by anti-gravitational repulsion, and so on.

This criticism is in effect is a time-reversibility rescue attempt for physics – a putative "CPTM theorem" wherein physics is invariant if we reverse *c*harges, mirror *p*arity-reflect, reverse *t*ime, *and* reverse the signs of all *m*asses. Although this criticism may sound promising, it is all either incorrect or confused and this rescue attempt fails, as we will now argue.

Consider plain Newtonian gravity. If we reverse the signs of all gravitational masses, then all forces Gm_1m_2/r^2 remain attractive. Thus circular orbits will remain solutions – both forwards and backwards in time – after reversing the signs of all gravitational masses. If we also reverse the sign of all inertial masses, then the accelerations induced by gravity will be (since $F = ma$) in the opposite direction to the forces, i.e. repulsive. In that case a putative "TM theorem" in Newtonian gravity (where "M" denotes changing the sign of *both* gravitational *and* inertial mass) would be *false* because circular orbits would no longer be solutions.

Conclusion: Newtonian mechanics is invariant under any subset of CPTM where M denotes reversing the signs of all gravitational masses but (and this is essential) *not* changing the signs of any inertial masses. Indeed more generally we could multiply all gravitational masses and all charges by j and multiply all inertial masses by j^2, where j is any (possibly complex) constant.

So the "rescue attempt" works in plain Newtonian mechanics. But it does not work in general relativity:

Theorem 3. *Reversing the sign of the mass-energy tensor T_α^β (thus reversing the signs of all mass-densities and pressures) does* not *lead to the same solutions (even time reversed) of Einstein's field equations, but instead to completely inequivalent metrics.*

Proof. As a simple example, consider the inequivalence of the "de Sitter" and "anti de Sitter" spaces [16] which have scalar curvatures of opposite signs, and utterly different topologies. Q.E.D.

Now let us counter all of the promising sounding "examples" of our critic.

The negative-mass Schwarzschild metric, while it is undoubtably a solution of the Einstein vacuum field equations everywhere that $r > 0$, sadly is *not* the time-reverse of a positive-mass Schwarzschild metric. For one thing, it no longer has any sort of horizon at $r = 2|M|$ and hence its geodesics behave quite differently – the central singularity of the negative-mass Schwarzschild metric is "naked," i.e. bidirectionally causally connected to spatial infinity. The negative-mass Schwarzschild metric with r *and* t *both* negated *is* the time-reverse of the positive-mass Schwarzschild metric... but that is merely because it *is* the positive-mass Schwarzschild metric (which is its own time-reverse)! The "white holes that only bifurcate" *are* the *same* metric as black holes that only merge. The gas (since its particles have positive inertial masses) still has positive pressure even if the signs of the gravitational masses of its particles are reversed, and those particles still would gravitationally attract. It is *not* the case that small perturbations of negative-mass Schwarzschild metrics will behave the same as the time reversals of small perturbations of positive-mass Schwarzschild metrics.

(3) We are now mentally prepared to make clear the genuinely time-irreversible and genuinely positive-mass-favoring nature of GR.

The Schwarzschild metrics are the only spherically symmetric solution of the Einstein vacuum field equations ("Birkhoff's theorem"). If we restrict ourselves to solutions without naked singularities bidirectionally causally connected to spatial infinity, then only the nonnegative-mass Schwarzschild metrics remain. The positive-mass Schwarzschild metrics all have event horizons which future-timelike geodesics enter but from which they cannot emerge. That defines a "direction of time." We have seen that small perturbations of the positive-mass Schwarzschild metrics have eternal solutions forward in time but generically have no solutions backward in time. The nonnegative mass Schwarzschild metrics are "attractors" – metrics nearby in suitable norms evolve toward them. *All these are genuine irreversibility and genuine positive-mass theorems* which do *not* depend upon any "input assumptions" about positive masses. Some might quibble that they do depend on the fact that the event horizons are inward=future oriented, but we reply that that is not an assumption, but rather a definition of which direction the "future" is. The fact that further phenomena (such as the perturbations-die claim or the fact that that holes can merge but not bifurcate) then agree self-consistently with that definition are genuine theorems *not* depending on making one new definition of "future" per new theorem. The properties of general relativity itself *output* this definition of "future."

(4) It is also possible to make versions of the "positive mass theorem" which are phrased in ways which avoid asymmetric energy condition assumptions. Here is such a theorem. (The statement we give here is an immediate consequence of theorem 1 on page 106 of [40]):

Theorem 4: *Let an asymptotically-flat static smooth spacetime be filled (perhaps at varying density ρ, and with the sign of ρ allowed to vary) with an isotropic gas obeying some equation of state $P(\rho)$, where P is the pressure, with the property that $c_1\rho \le P \le c_2\rho$ for some positive constants c_1 and c_2. Then: the total ADM mass of that spacetime is non-negative.*

Note in theorem 4 that no assumption was made about the sign of ρ. Also, there are various ways [11] [12] [13] [17] in which flat-space vacuum quantum field theories favor positive mass. In all of these cases the preferred-positivity of mass arises as an *output* without any inputted sign-assumptions, and our "critic" is answered.

6 An analogy in 3D Euclidean geometry

Some readers have still felt they could not understand how it could be logically possible that GR, a theory which infinitesimally *locally* has time-reversal invariance, somehow through nonlocal effects[8] engenders time-irreversibilities.

To help them, we now describe a similar phenomenon in 3D Euclidean geometry (an area hopefully far easier to understand than GR!). Consider the theory of 2D finite-area boundaryless surfaces in 3D. The statement that such a surface is *convex* is expressible purely locally, e.g. as a statement that it everywhere has positive curvature[9], i.e. that small triangles drawn on the surface have angle sums exceeding π.

Now, everybody knows the theorem that any such convex body has an "inside" (of finite 3-volume) and an "outside" (of infinite volume). How can it be that a local postulate, completely symmetric with respect to "side" of the surface, has led to this asymmetric consequence – favoring one side over the other?

You may prefer not to regard this as a "theorem" but rather as the *definition* of the word "inside." The riposte then is the genuine asymmetric theorem that the intersection of two convex body *in*teriors is another convex body interior, whereas the intersection of two convex body *ex*teriors is (in general) *not* the exterior of any convex body[10]. Please direct any further complaints to Euclid!

7 Time-arrow Manifesto

The overenthusiastic incorrect claim is often made by physicists that "all of the known laws of physics are invariant under CPT (reverse all charges, perform a mirror reflection, and reverse time)."

In fact there are two counterexamples to that claim:

1. "measurement" in quantum mechanics, and
2. numerous theorems in general relativity (discussed above, some new).

The truth: What is commonly called the "CPT theorem" only applies to (a wide class of) quantum field theories in *flat* spacetime (i.e. no gravity allowed!) and in which quantum measurement is *forbidden*. Further: Previous claims that General Relativity is time-reversal invariant are in fact false or at least misleading as we have shown.

Manifesto: *These two phenomena are the same thing! The ultimate logical source of all time-irreversibility is gravity!*[11]

[8] The Schwarzschild diameter of a $7M_{\rm sun}$ black hole is 41km, the size of a large city. While nonlocal compared to infinitesimal and Planck scales, this is still very small compared to, e.g. the Earth.

[9] Actually, a sphere with a spherical inward "dimple" has positive *intrinsic* curvature everywhere but is nonconvex. However, that could be forbidden by either demanding smoothness of the 3D embedding, or by maximizing the minimum 3-volume on either side of the surface over all embeddings, smooth or not. It is then a theorem of Aleksandrov and Pogorelov [37] that a convex surface metric is embeddable in 3-space in a *unique* way (up to congruence), i.e. "convex shells are rigid."

[10] This is intentionally analogous to not regarding the fact that a positive-mass Schwarzschild black hole has a future-inward horizon as a theorem, but rather as a *definition* of the word "future," in which case a genuine theorem is that black holes can increase but cannot decrease their surface area.

[11] I claim that the whole issue of the arrow of time vis a vis gravitational physics has been misunderstood and/or shallowly treated in the usual GR books. E.g. Weinberg ([47] p.597) confines his discussion to raising the (wrong) notion that if the universe "turned around," i.e. began to contract, the 2nd law of thermodynamics would break down and time's arrow would vanish. To his credit Weinberg casts doubt on that, but it later was actually stated as a fact by Hawking in his popular book "A brief history of time." (The "justification" for this is supposedly that the expanding universe provides a heat sink allowing heat engines, including life forms, to operate and thus experience "perceptual-thermodynamic time." This is totally wrong: Actually if the universe started contracting

8 Quantum measurement and consequences

The Copenhagen interpretation of quantum mechanics involves two things: (a) time-evolution via Schrödinger's equation (which may be regarded as a deterministic unitary transformation in a Hilbert space) and (b) "measurement" (which involves a nondeterministic sampling, and also is a nonunitary operation).

Unfortunately, it was never clear just what a "measurement" is. Who decides when something is being "measured"? Why do measurements affect wavefunctions instantly without being affected by speed of light limitations?

Every theorist has always been very unhappy about measurement[12], and indeed the approach of mainstream physics nowadays is to deny the existence of Copenhagen measurement, claim that only Schrödinger time evolution happens, and claim that measurement somehow effectively get simulated. But I claim that all previous attempts to get quantum "measurement without measurement" are fundamentally flawed or at least highly debatable and nonrigorous; but the Manifesto of the present paper shows how to get measurement-like effects in a undebatable way.

When Von Neumann attempted to make all these things mathematically precise, he introduced his so-called "density matrix formalism" and a quantity he called "entropy." He proved

Theorem 5: *In Von Neumann's density matrix formalization of quantum mechanics, Schrödinger time-evolution is reversible and leaves entropy unchanged, but any measurement increases entropy and is irreversible.*

(Well known) Proof: Von Neumann's "density matrix" ρ, which describes the state of the system, is by construction a nonnegative definite Hermitian operator whose real eigenvalues sum to 1. In other words the spectrum of ρ forms a probability distribution. The "entropy" is the entropy of[13] that probability distribution. Let H be the (Hermitian) Hamiltonian operator and let t be time. "Schrödinger time evolution" $\dot{\rho} = \frac{i}{\hbar}(\rho H - H\rho)$ merely performs a unitary similarity transformation $\rho \to Q\rho Q^{-1}$ where $Q = e^{iHt/\hbar}$ is unitary (i.e. Q^{-1} is the Hermitian adjoint of Q). Any similarity transformation leaves the spectrum of ρ, and hence S, unchanged. Any unitary transformation leaves the Frobenius norm (sum of the squared moduli of the entries of ρ) unchanged.

"Measurement" zeros the off-diagonal elements of some block of the density matrix (in some basis). This converts "pure states" (rank-1 density matrices) to "mixed states" (rank>1) whereas unitary time evolution preserves rank; this proves measurement is not unitary. Measurement always preserves nonnegative definiteness, trace, and Hermiticity, and hence the fact that ρ's spectrum is a probability distribution. But it is irreversible (since information is destroyed by the zeroing – there is in general no way to back-deduce the pre-measurement ρ from the post-measurement ρ) and it plainly decreases ρ's Frobenius norm (unless it is an ineffectual measurement leaving ρ unchanged). The fact that measurement always increases entropy is harder. It depends on the fact[14] ([35] propositions 1.6 & 3.1) that entropy is a convex (i.e. concave-$\cap$) function of Hermitian matrices with eigenvalues in $[0,1]$, i.e. $\mu S(A) + (1-\mu)S(B) \leq S(\mu A + [1-\mu]B)$ if $0 \leq \mu \leq 1$ and $\lambda(A), \lambda(B) \subseteq [0,1]$. We now regard measurement as replacing ρ by a convex combination of various unitary similarity transformations of ρ (each of which rotates the complex phase angle of some off-diagonal element of ρ); the fact that measurement increases S now immediately follows. Q.E.D.

This is different from classical mechanics. In classical mechanics, entropy, regarded as a phase space volume, never can change, although certain notions of entropy (regarded by observers with only partial, statistical descriptions of the system) can change, and usually increase. However, there is always a tiny probability this entropy will decrease drastically, e.g. that all the gas molecules will decide to fly to the West side of the room. Indeed "Poincare's recurrence theorem" says that, if one waits long enough, eventually every state of a bound classical dynamical system with a finite number of degrees of freedom will be revisited arbitrarily closely. Thus in classical mechanics, entropy decreases as much as it increases. (Boltzmann was said to have responded "you should live that long.") Meanwhile, in quantum mechanics, Von Neumann's entropy really *does* increase and really never can decrease, so there is no recurrence theorem – but in the *absence* of measurement, entropy never increases, and in any compact state space recurrence is forced.

tomorrow, the stars still would shine and the solar system would remain hotter than the rest of space for many billions of years, allowing us to continue living on, happily running our heat engines. By switching to accretion disks around black holes as an energy source, heat engines could continue to run until the universe became hotter than Xrays. Meanwhile even in an eternally expanding universe the stars and other heat sources ultimately would burn out and heat engines would no longer be able to make progress. I claim that the geometric arrow of time and thermodynamic arrow of time will always be in the same direction, and entropy will always continue to increase; the worst that can happen is that entropy increase could ultimately become very slow, preventing human life. In that case from some human observer point of view time might be regarded no longer as existing, but it definitely will not turn backwards!) Wald [46] (and similarly [30] p.922-) confines his discussion to the fact that oriented $(3+1)$-manifolds have a "future" and a "past" but gives no a priori definition of which is which, apparently regarding that as an arbitrary "extra input" to GR rather than (as it in fact is) an output. There have been many authors who have approvingly mentioned both (1) the alleged time-reversibility of GR and (2) the plainly-irreversible "laws of black hole mechanics," never noticing the apparent contradiction between the two. Although certainly it was well known that many such theorems depend on "energy conditions" [16], I have never seen an explicit statement saying the arrow of time is due to positivity of mass.

[12] But it does not bother the experimentalists because their notions of what a "measurement" is – mathematically-imprecise though they may be – always seem to work!

[13] The entropy S of a probability distribution p_k is $S \equiv -\sum_k p_k \log p_k$.

[14] Actually, even more strongly, the entropy function $x \to -x \log x$ is a "convex matrix function on $H_n(0,1)$" in the notation of Horn and Johnson ([19] chapter 6.6).

Numerous anti-Copenhagen theorists (nowadays the mainstream) proposed that the effect of measurement is *simulated* by time evolution of a quantum system with a large number of degrees of freedom, as far as beings who "mostly have fewer degrees of freedom" perceive it. A unitary transformation on large-dimensional density matrices will usually appear, to those of us who can only see a small subset of those dimensions ("our part of the universe") and regard the others (the "outside environment") as "random noise," to act in a non-unitary, nondeterministic, usually-information-losing way. Hopefully, whenever this happens, the off-diagonal terms in our submatrix ("off-diagonal" in the position basis) get their complex phase angles randomized. That has essentially the same effect as a Von Neumann non-unitary position measurement operation.

But that whole approach has never been rigorously justified. What if there *is* no larger system and no external environment since we are modeling "the wave function of the whole universe"? Why is the "outside environment" necessarily "random" and uncorrelated with the system we are concerned with?[15] How does the "classical limit" arise, exactly? How can notions of "what we consciously experience" be made mathematically precise? How did the position basis somehow become "favored" (the *positions* of objects keep getting effectively "measured" while other quantities remain unmeasured) – despite the fact that all orthonormal bases of a Hilbert space are unitarily equivalent, and despite the possibility that in the early universe there was no such favoritism? Why do we never experience "quantum weirdness" such as being in a superposition of being in Paris and Tokyo, and why can't things keep getting weirder?

Although there have been attempts to answer all of these questions, I think it is safe to say that no explanation has answered all of them simultaneously in a mathematically rigorous and physically satisfying way, and there is no hope of that happening in the forseeable future[16]. Furthermore, on the other side of this fight there *is* a physically satisfying and rigorous theorem [44], the "CPT theorem," giving a way in which quantum mechanics is time reversible, and theorem 5 proves that measurement-free quantum mechanics can never increase entropy and can never do anything irreversible. It is for that reason that the manifesto says that gravity is the *only* source of time-irreversibility in physics, since all non-gravitational physics[17] is quantum-mechanical, and hence reversible.

We advocate cutting this Gordian knot by simply proposing that

1. The laws of physics really *are* time-irreversible and position-based.
2. *Gravity* is the fundamental source of both the time-irreversibility and the primacy of position. Because gravity is the weakest force (by about 40 orders of magnitude) these effects are weak[18], but they nevertheless are nonzero.
3. In particular, gravitational phenomena are responsible for what are commonly called "decoherence," and "quantum measurement," and hence are the fundamental source of the "arrow of time," the "classical, nonquantum appearance of nature," the "lack of quantum weirdness in human experience," and the experimental validity of "outgoing-only radiative boundary conditions" (a.k.a. the preference for "retarded" rather than "advanced" time Green's functions) and of the "second law of thermodynamics."
4. The mystery of how the universe could initially have been in a very simple low entropy state (which, a priori, would seem unlikely) too is explained.

How do all the latter follow from the former? Clearly the whole notion of "metric" underlying GR gives primacy to position, and GR has, as we've seen, built-in time irreversibility properties and properties similar to outgoing-only radiative boundary conditions (including for electromagnetic waves, according to the argument at the end of §2). We now discuss[19] quantitatively a way that gravitational effects can lead to "decoherence" and "continual position measurement"; these in turn cause everything else on our list except for the final item about the early universe.

Suppose a mass M is orbiting the sun at radius $r \approx 1$ AU. It will emit gravitational waves of frequency $f = 2\text{year}^{-1}$ and thus gradually lose energy and spiral into the sun. The power P emitted is given by Einstein's formula (EQ 10.5.25 p.272 of [47])

$$P \approx M^2 r^4 f^6 G c^{-5}/10 \approx 9 \times 10^{-53} M^2 \text{watts/kg}^2. \quad (12)$$

Suppose that gravitational wave excitations obey the usual quantum energy relation $E = hf$. Then each emitted "graviton" will have energy $E = hf \approx 4 \times 10^{-41}$ Joules. Since each graviton emission represents "permanently lost information," it may be regarded as, effectively, a quantum measurement. Thus our mass M automatically has its position (or something very much like position) "measured" about once every $10hc^5 G^{-1} f^{-5} r^{-4} M^{-2} \approx 4 \times 10^{11} M^{-2}\text{kg}^2$seconds, accurate to about a graviton-wavelength. Thus the position of a 2000kg mass (a truck) in a 1 AU orbit around the sun would automatically be "measured" by this effect about once per day to an accuracy of order 1/2 lightyear. This

[15] See footnote 21.

[16] Because any physically satisfying environment clearly *does* become entangled with our system forever after the first interaction, and nobody has been able to formulate and prove the perpetual existence of any notion of "effective" lack of entanglement.

[17] The "standard model," that is.

[18] This weakness is a fortunate thing since all biochemistry depends heavily on "microscopic reversibility" to allow metabolism to be energy-efficient. If time-reversibility were strongly violated, life as we know it could not exist.

[19] See also [41] for a different mechanism of "gravitational decoherence."

certainly seems a very small (but nonzero) effect![20] On the other hand, a 1AU black hole binary (each hole 10 solar masses) would be measured by this effect about 10^{51} times per second accurate to, say 1/6 lightyear. If all the measurements/emissions are regarded as independent then n such measurements would yield a positional accuracy of about $n^{-1/2}/6$ lightyears. That would mean that each black hole would have its position measured to an accuracy of about 10^{-14} meters once each year purely by this effect. Since this measurement process would be continual (the "quantum watchdog effect"), it would be impossible for either black hole ever to acquire much positional uncertainty.

Now consider cosmic rays, photons, or neutrinos flying by our black holes. Each time the cosmic ray either fell in to the hole or did not, that could be regarded as an irreversible measurement of the cosmic ray's location by the hole. Now, these cosmic rays could then fly onward and measure things about the rest of the universe (or could be regarded as having measured things about the rest of the universe previously). A classic model of environmental decoherence [21] is that photons, cosmic rays, neutrinos, etc, that fly by and interact with you act to continually measure your position. More precisely, an interaction potential V for a timespan t leads to a phase-angle rotation of order $Vt/\hbar$; if Vt depends on your position, the net effect of this is to randomize phase angles of off-diagonal terms in your density matrix in the position-basis. This explains why you never experience being in a superposition of being located in Paris and Tokyo. Numerical estimates [45] of the known cross sections for interactions between you and neutrinos, cosmic rays, air molecules, photons, etc., and the known fluxes of these things indicate that the rates of "measurement" of your position (even from neutrinos) are *enormous*. But what those papers did not say is: this only works if those particles themselves can be regarded as localized, as opposed to, themselves being in some quantum superposition of being in Paris and Tokyo (in which case an interaction with you would *not* measure your position)[21]. The complete fallacy of the particle flyby model can be revealed as follows. Suppose you, neutrinos, and everything else in the universe were in maximally *de*localized states, i.e. plane waves, i.e. momentum eigenstates. Since collision cross sections between you and neutrinos are highly sensitive to your relative momenta, an interaction between you and a neutrino would then not measure your position at all, but might provide quite a good constraint on your momentum. In that case your *momentum* would presumably be being continually "measured," so that the effect of such "environmental decoherence" would be to continually make our momenta more certain but consequently our positions *less* certain! That kind of universe seems entirely different.

So to make the particle-flyby position decoherence model work, and to allow the position basis to be favored, at some point in the cascade of measurements of things by other things, there has to be a "ground truth" – something whose position really *is* highly certain. We've seen that such a ground truth is provided in our universe by any very large orbiting masses. This is not necessarily the only source of ground truth, but it is logically sufficient. Even a very small rate of input of ground truth should have remarkably powerful contagious effects. Thus just one cosmic ray can measure the location of the entire sun, and subsequently photons from the sun can measure the location of everything in the solar system, other stars, etc.

In this picture, we now see that what people call "increase in entropy" and "measurement" really is a unidirectional[22] *transfer* of information, also known as heat, from ordinary matter such as ourselves to ripples in the metric of spacetime. Ordinary matter gradually becomes more measured and loses heat.

It is no longer a "great mystery" why the universe was initially created in a low-entropy (i.e., ultra-unlikely) state, permitting entropy to continue to increase thereafter. In fact, the universe could have been created in a high entropy state – and entropy (of matter) would *still* continue to increase thereafter. All that matters is that the flow of information between matter and gravity-

[20] Also, charges on our orbiting mass, since accelerated, or merely because they are warm, would emit electromagnetic radiation. And, of course, the sun also emits light and neutrinos. By our arguments at the end of §2, almost all of this radiation too will not "come back" to ordinary matter, and hence also should lead to decoherence effects – and usually much larger ones.

[21] If you think it is implausible that incoming neutrinos would somehow "know" to have precisely the right wavefunction to avoid measuring your position, then consider the fact that, in time-reversed quantum mechanics, precisely this sort of thing *is* assumed to happen. "Anti-measurement" like time-evolution effects must happen in time-reversed quantum mechanics if "measurement-like" evolution effects happen in the forward direction. The latter is multiplying off-diagonal density metric elements by random phase factors; the former is multiplying them by very *non*random phase factors. Thus it really is, mathematically, exactly as implausible that measurement (corresponding to our physical intuition) happens as that anti-measurement (completely violating our physical intuition) happens. Now since, by the CPT theorem, quantum mechanical time evolution acts the same in both time directions, we see the complete fallacy of "obviously, incoming things act randomly" arguments to "explain" the arrow of time. If there is to be any such explanation purely within nongravitational quantum mechanics, then by CPT it must involve the facts that, e.g., the universe is mostly matter and not antimatter, and neutrinos are mostly lefthanded and not righthanded – but current books [15] about decoherence never regard that as its root cause. Even if the prevalence of matter over antimatter *were* somehow regarded as responsible for time's arrow, then we would have no such explanation in a universe with equal amounts of the two – except for our, gravitational, explanation.

[22] Actually, it is misleading for me to use the word "unidirectional." Some matter that emits a graviton can be said to have "transferred information" (concerning angular momentum, say) to the land of gravitons, but we may equally well regard this as a transfer of the same amount of information in the other direction. There is directionality in the flow of energy here, but really the flow of information is bidirectional. Both sides regard themselves as having just seen a "random" effect, i.e. of just having acquired some "random bits" from the other side. The key point is that such random bits *stay* random, i.e. are statistically independent of both previously transmitted and future random bits, in the sense that gravitons cannot "come back."

ripples is *one way* to allow subsequent "quantum measurement" and "2nd law" effects to happen.

9 Consequences for theories of quantum gravity

S.W.Hawking pointed out that Hawking radiation evaporation of black holes (which happens in certain minimal quantum extensions of classical GR) seems necessarily to lead to non-unitary "information loss." (You toss an encylopedia into a black hole. Completely random thermal noise ultimately comes out[23].) Pure states tossed into a black hole come out as mixed states. I.e., Hawking evaporation is, in this sense, a quantum measurement effect. Hawking also observed that this must also lead to nonconservation of baryon and lepton number.

The horrified reaction to this by other physicists was a sight to see. J.Preskill in his excellent review article [38] on this even concluded that "the most conservative hypothesis" needed to try to wriggle out of this trap was that black hole creation, as a side effect, creates disconnected "baby universes" to absorb the lost information.

Let us now reexamine this armed with our new perspective. We now know how "measurement" can happen, even if the wavefunction of the whole universe is being modeled. It is no longer a horror that must be expunged from physics at all cost, but rather, a natural consequence of gravity.

A substantial fraction of Hawking radiation is gravitational and hence, even if Hawking's process *were* fully unitary, that information *still* would, by our theorem 1, be irretrievably "lost" to those of us made of ordinary matter. So there is clearly a lot of "measurement" and "information loss" here. We now see that the Hawking information-loss process is *another* way in which black holes – now even *non*orbiting holes – provide a source of "ground truth" needed to make the whole decoherence bandwagon [15] proceed. This is not bad. This is good.

Still, though, we are left with the puzzle of how to put non-unitarity into quantum physics. At first I thought that Hawking's paradox and the irreversibility properties in the present paper would force any marriage of quantum mechanics and gravity to be nonunitary. However, my current view is that it is permissible to keep it pure-unitary provided there is a reservoir of extra Hilbert space dimensions (a.k.a. extra dynamical degrees of freedom) with which

1. "our" degrees of freedom can occasionally interact *once* and (almost always) *only* once so that "lost" information cannot "come back" from them and
2. those interactions must be highly *position*-sensitive so that the position basis becomes the favored one.

That course gives us the best of both worlds: we enjoy the benefits of unitarity while at the same time permitting irreversible effects such as entropy increase and measurement needed to make theoretical physics resemble reality. We have already seen in classical GR that gravity wave ripples constitute such a reservoir[24].

[23] Merely burning the encyclopedia would not in principle destroy its information, since in principle the encyclopedia could be reconstructed from complete knowledge of the quantum states of the combustion products. The horror of Hawking's process is that it *genuinely* seems irreversible.

[24] In future work I plan to expound a new theory of quantum gravity I call "QGN." It indeed involves such a reservoir.

References

[1] N. Andersson: On the asymptotic distribution of quasinormal-mode frequencies for Schwarzschild black holes, Class. Quantum Grav. 10 (1993) L61-L67.

[2] N. Andersson & C.J.Howls: The asymptotic quasinormal mode spectrum of non-rotating black holes, `gr-qc/0307020`.

[3] N. Andersson, M.E.Araujo, B.F.Schutz: Generalized Bohr-Sommerfeld formula for the Schwarzschild black hole normal modes, Class. Quantum Grav. 10,4 (1993) 757-765.

[4] A. Bachelot & A. Motet-Bachelot: Les resonances d'un trou noir de Schwarzschild, Ann. Inst. Henri Poincare 59 (1993) 3-.

[5] J.M.Bardeen, B.Carter, S.W.Hawking: The four laws of black hole mechanics, Commun.Math.Phys. 31,2 (1973) 161-170.

[6] A.S.Barreto & M.Zworski: Distribution of resonances for spherical black holes, Math. Res. Lett. 4 (1997) 103-121.

[7] H-J. Blome, B.Mashhon: Quasinormal oscillations of a Schwarzschild black hole, Phys.Lett. A 100,5 (1984) 231-234.

[8] S.Chandrasekhar: The Mathematical Theory of Black Holes, Clarendon Press, Oxford 1983. Also: The mathematical theory of black holes and of colliding plane waves, University of Chicago Press 1991.

[9] Yvonne Choquet-Bruhat wrote her first big paper on The Cauchy problem for General Relativity in 1952 and in 2001 reviewed contributions by her and others to this problem in `gr-qc/0111017`.

[10] D.M.Christodolou & S.Klainerman: The global nonlinear stability of Minkowski space, Princeton Univ. Press 1993.

[11] C.J.Fewster & Edward Teo: Bounds on negative energy densities in static spacetimes, Phys.Rev.D 59 (1999) 104016

[12] C.J.Fewster & Edward Teo: Quantum inequalities and quantum interest as eigenvalue problems, Phys.Rev.D 61 (2000) 084012

[13] L.H.Ford & T.A.Roman: Averaged Energy Conditions and Quantum Inequalities, Phys. Rev. D 51 (1996) 4277-4286. gr-qc/9410043

[14] D.Garfinkle & R.M.Wald: On the possibility of a box for holding gravitational radiation in thermal equilibrium, Gen. Relat. Gravit. 17,5 (1984) 461-473.

[15] D.Guilini, E.Joos, C.Kiefer, J.Kupsch, I-O.Stametescu, H.D.Zeh: Decoherence and the appearance of a classical world in quantum theory, Springer 1996.

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

[17] Adam D. Helfer: Operational energy conditions, Classical & Quantum Gravity 15,5 (1998) 1169-1183.

[18] G. 't Hooft: Introduction to General relativity, Rinton Press, Inc., Princeton NJ, 2001.

[19] R.A.Horn & C.R.Johnson: Topics in matrix analysis, Cambridge Univ. Press 1991.

[20] Fritz John: Partial differential equations, Springer (Applied math'l sciences #1) 4th ed. 1982.

[21] E.Joos & D.Zeh: Zeitschrift Phys. B 59 (1985) 223-

[22] S.Klainerman & F.Nicolo: The evolution problem in general relativity, Birkhauser (Progress in math'l physics #25) 2003.

[23] Kostas
D. Kokkotas & Bernd G. Schmidt: Quasi-Normal Modes of Stars and Black Holes, Living Reviews in Relativity 2 (1999). [This is an electronic journal, available at `http://www.livingreviews.org`.]

[24] D.Kramer, H.Stephani, E.Herlt, M.MacCallum: Exact solutions of Einstein's field equations, Cambridge Univ. Press & VEB Deutscher Verlag 1980. There also is a 2003 Cambridge Univ. Press 2nd edition.

[25] Edward W. Leaver: An analytic representation for the quasi-normal modes of Kerr black holes, Proc. Royal Soc. London A 402 (1985) 285-298

[26] Edward W. Leaver: Spectral decomposition of the perturbation response of the Schwarzschild geometry, Phys. Rev. D 34 (1986) 384-408

[27] Edward W. Leaver: Remarks on the continued fraction method for computing black hole quasinormal frequencies and modes, Phys. Rev. D 45 (1992) 4713-4716.

[28] H.Liu: Asymptotic behaviour of quasi-normal modes of Schwarzschild black holes, Class. Quantum Grav. 12 (1995) 543-552.

[29] H.Liu & B. Mashhoon: On the spectrum of oscillations of a Schwarzschild black hole, Class. Quantum Grav. 13 (1996) 233-251.

[30] C.W.Misner, K.S.Thorne, J.A.Wheeler: Gravitation, Freeman 1973.

[31] Lubos Motl: An analytical calculation of asymptotic Schwarzschild quasinormal frequencies, Adv.Theor.Math.Phys. 6 (2003) 1135-1162; `gr-qc/0212096`.

[32] Lubos Motl & Andrew Neitzke: Asymptotic black hole
quasinormal frequencies, Adv.Theor.Math.Phys. 7 (2003) 307-330; `gr-qc/0301173`;

[33] H-P. Nollert: Quasinormal modes of Schwarzschild black holes: The determination of quasinormal frequencies with very large imaginary parts, Phys. Rev. D 47 (1993) 5253-5258.

[34] H-P. Nollert: Quasinormal modes the characteristic 'sound' of black holes and neutron stars, Classical Quantum Gravity 16 (1999) R159-. (Review article.)

[35] M.Ohya & D.Petz: Quantum entropy and its use, Springer 1993.

[36] R.Penrose, R.D.Sorkin, E.Woolgar: A positive mass theorem based on the focusing and retardation of null geodesics, gr-qc/9301015. [submitted Commun. Math. Phys.; to appear in Phys. Rev. D 60 (1999)?]

[37] A.V.Pogorelov: Extrinsic Geometry of convex surfaces, AMS 1973 (Translations of math monogrs #35).

[38] John Preskill: Do black holes destroy information?, Int'l symposium on black holes, membranes, wormholes, and superstrings, The Woodlands, Texas (16-18 Jan 1992). `hep-th/9209058`.

[39] T.Regge & J.A.Wheeler: Stability of a Schwarzschild singularity, Phys. Rev. 108 (1957) 1063-1069.

[40] T.Shiromizu & M.Sugai: Local energy conditions and asymptotic conditions, Classical & Quantum Gravity 11,7 (1994) L103-L108.

[41] Warren D. Smith: An argument against quantum computers, 2003. Available from `http://www.math.temple.edu/~wds/homepage/works.htm`

[42] L.Smolin: The thermodynamics of gravitational radiation, General Relativity and Gravitation 16,6 (1984) 205-210.

[43] L.Smolin: On the intrinsic entropy of the gravitational field, General Relativity and Gravitation 17,5 (1985) 417-437.

[44] R.F. Streater and A.S. Wightman, PCT, Spin & Statistics, and All That, Benjamin 1968.

[45] Max Tegmark: Apparent wave function collapse caused by scattering, Found. Phys. Letters 6 (1993) 571-590.

[46] Robert M. Wald: General relativity, University of Chicago press 1984.

[47] Steven Weinberg: Gravitation and cosmology, Wiley 1972.

[48] E.Witten: A new proof of the positive energy theorem, Commun. Math. Phys. 80 (1981) 381-402. Followup papers correcting and extending Witten's proof are: O.Reula: Existence theorem for solutions of Witten's equations and nonnegativity of total energy, J.Math.Phys. 23 (1982) 810-814; T.Parker & C.H.Taubes: On Witten's proof of the positive energy theorem, Commun.Math.Phys. 84 (1982) 223-238; J.M.Nester: A new gravitational energy expression with a simple positive energy proof, Phys.Lett 83A (1981) 241-242; M.Ludvigsen & J.A.G.Vickers: A simple proof of the positivity of the Bondi mass, J.Phys.A15 (1981) L67-L70; G.W.Gibbons & S.W.Hawking: A Bogomolny bound for General Relativity and Solitons in $N = 2$ supersymmetry, Phys.Lett B 109 (1982) 190-194; G.W.Gibbons, S.W.Hawking, G.T.Horowitz, M.J.Perry: Commun.Math'l. Phys. 88 (1983) 295-308. M.Herzlich: Positive mass theorem for black holes, J. Geometry & Physics 26,1-2 (1998) 97-111. These show that the "mass" (as measured at spatial infinity – this is called the "ADM mass") of any smooth $(3+1)$-dimensional spacetime metric (assuming the TCC) is non-negative. The extensions show this non-negativity is also true if the mass is measured at future-null infinity (this is called the "Bondi mass," and under the Einstein equations it can decrease with time due to gravitational radiation; the ADM mass is always constant); Gibbons and Hawking's paper extended this to hold under the Einstein-*Maxwell* equations for sufficiently weakly-*charged* matter (obeying a generalization of the DEC), the last two papers extended it to hold for spacetimes including black holes (i.e., an event horizon, with anything whatsoever on the unseeable side of that horizon). A different and less simple proof than Witten's is: R.Schoen & S-T.Yau: Commun. Math. Phys. 65 (1979) 45-76; 79,1 (1981) 47-51; 79,2 (1981) 231-260; Phys.Rev.Lett. 48,6 (1982) 369-371. See also: G.Bergqvist: Simplified spinorial proof of the positive energy theorem, Phys.Rev.D 48,2 (1993) 628-630, and [36] where the required energy-condition assumptions are weakened to the ANEC (averaged null energy condition) only.