

This is a repository copy of *A semiclassical singularity theorem*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/183750/>

Version: Published Version

Article:

Fewster, Chris orcid.org/0000-0001-8915-5321 and Kontou, Eleni orcid.org/0000-0003-4409-8188 (2022) A semiclassical singularity theorem. *Classical and Quantum Gravity*. 075028. ISSN: 1361-6382

<https://doi.org/10.1088/1361-6382/ac566b>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

PAPER • OPEN ACCESS

A semiclassical singularity theorem

To cite this article: Christopher J Fewster and Eleni-Alexandra Kontou 2022 *Class. Quantum Grav.* **39** 075028

View the [article online](#) for updates and enhancements.

You may also like

- [Did the universe experienced a pressure non-crushing type cosmological singularity in the recent past?](#)
S.D. Odintsov and Vasilis K. Oikonomou
- [Perspective—An Age of Sensors](#)
Anthony Turner
- [Doubly Charged Higgs Production at Future *ep* Colliders](#)
Xing-hua Yang and Zhong-Juan Yang



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

A semiclassical singularity theorem

Christopher J Fewster¹  and
Eleni-Alexandra Kontou^{1,2,3,*} 

¹ Department of Mathematics, University of York, Heslington, York YO10 5DD, United Kingdom

² Department of Physics, College of the Holy Cross, Worcester, MA 01610, United States of America

³ ITFA and GRAPPA, Universiteit van Amsterdam, Science Park 904, Amsterdam, The Netherlands

E-mail: e.a.kontou@uva.nl

Received 27 October 2021, revised 3 February 2022

Accepted for publication 17 February 2022

Published 18 March 2022



Abstract

Quantum fields do not satisfy the pointwise energy conditions that are assumed in the original singularity theorems of Penrose and Hawking. Accordingly, semiclassical quantum gravity lies outside their scope. Although a number of singularity theorems have been derived under weakened energy conditions, none is directly derived from quantum field theory. Here, we employ a quantum energy inequality satisfied by the quantized minimally coupled linear scalar field to derive a singularity theorem valid in semiclassical gravity. By considering a toy cosmological model, we show that our result predicts timelike geodesic incompleteness on plausible timescales with reasonable conditions at a spacelike Cauchy surface.

Keywords: energy conditions, semiclassical gravity, singularity theorem, quantum energy inequalities

(Some figures may appear in colour only in the online journal)

1. Introduction

Singularity theorems represent an effort to answer a central question in gravitational physics: under which conditions do cosmological or astrophysical systems originate or end in a singularity? In effect: are singularities inevitable in our Universe? The famous singularity theorems of Penrose [1] and Hawking [2] address this question in general relativity by taking geodesic

* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

incompleteness as the defining characteristic of a singularity, and showing that this occurs, from suitable starting situations, for any matter model obeying energy conditions broadly expressing the idea of local energy positivity. In particular, Penrose’s theorem on null geodesic incompleteness was based on the null energy condition (NEC), that the stress–energy tensor should obey $T_{\mu\nu}\ell^\mu\ell^\nu \geq 0$ for all null vectors ℓ^μ , while Hawking’s result on timelike geodesic incompleteness required the strong energy condition (SEC), that

$$\rho_U := T_{\mu\nu}U^\mu U^\nu - \frac{T^\mu{}_\mu}{n-2} \geq 0 \quad \text{for all unit timelike vectors } U, \quad (1)$$

where n is the spacetime dimension. Here, ρ_U is the *effective energy density* (EED) seen by an observer with velocity U^μ (cf reference [3], which refers to the ‘effective density of gravitational mass’). Via the Einstein equation, the SEC is equivalent to the geometrical condition $R_{\mu\nu}U^\mu U^\nu \leq 0$ on the Ricci tensor (our conventions are listed at the end of the introduction).

It has long been recognized that the energy conditions are too restrictive. Even the classical massive minimally coupled scalar field violates the SEC and the non-minimally coupled scalar field violates the NEC, while quantum fields violate all pointwise energy conditions [4]. Reviews concerning the status of energy conditions include references [5, 6], while reference [7] is a comprehensive review of all aspects of singularity theorems. Accordingly, several authors have focused on relaxing the energy condition in singularity theorems looking instead at weaker, averaged energy conditions [8–11]. Most of that work focuses on Penrose-type (null geodesic incompleteness) results and the half-averaged NEC. More recently, reference [12] presented proofs of both Hawking-type (timelike geodesic incompleteness) and Penrose-type singularity theorems with weakened energy conditions. Based on their results, reference [13] proved a Hawking-type singularity theorem for the classical Einstein–Klein–Gordon field. The essential technical ingredient in this work, as with almost all of the previous literature concerning weakened energy conditions, is the analysis of a Riccati differential inequality derived from the Raychaudhuri equations [14]. However, it has been realised very recently [15] that *index form* methods provide a much simpler route to obtaining singularity theorems with weakened energy hypotheses, with the advantage that they also provide quantitative estimates of the timescale on which the singularity occurs. Index form methods appear in some textbook accounts of the original singularity theorems [16, 17] and are also employed in generalizations of the singularity theorems (and related results) in other directions—see e.g. [18–20]. An interesting account of the interplay between Riccati and index-form methods in a slightly different context can be found in reference [21].

But what about quantum fields? Reconciliation of the energy conditions and quantum field theory began with the seminal works of Ford [22, 23]. Motivated by an apparent conflict between negative energy densities and the second law of thermodynamics, he suggested and later proved that quantum field theory contains mechanisms that impose fundamental restrictions on the possible magnitude and duration of any negative energy densities or fluxes due to the quantum field; these restrictions—*quantum energy inequalities* (QEIs)—have been proved to hold for a range of different quantum field theories in flat and curved spacetimes. See [6, 24] for recent reviews.

An important question is whether QEIs can be used as hypotheses for singularity theorems and this has been addressed to some extent. Both references [12, 15] introduced singularity theorems with hypotheses inspired by QEIs. For example, reference [15] proved a Hawking-type singularity theorem in spacetime dimension $n = 2m$ ($m \in \mathbb{N}$) under the following geometrical condition on the Ricci tensor: for any unit-speed timelike geodesic $\gamma : I \rightarrow M$, where I is a compact interval in \mathbb{R} , it is assumed that there are non-negative constants $Q_m(\gamma)$ and $Q_0(\gamma)$ of appropriate dimensions so that the inequality

$$\int_I f(t)^2 R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu |_{\gamma(t)} dt \leq Q_m(\gamma) \|f^{(m)}\|^2 + Q_0(\gamma) \|f\|^2, \quad (2)$$

holds for all smooth real-valued f supported in the interior of I . Here $\|\cdot\|$ denotes the standard norm of $L^2(I)$. Under this condition, and additional assumptions on the extrinsic curvature of a spacelike Cauchy surface, it was shown that timelike geodesic completeness must fail. Quantitative estimates of the required initial contraction and timescales for the singularity were obtained in various cases, including models drawn from cosmology.

However, none of the conditions used in the works just mentioned were derived directly from quantum field theories. Indeed, QEIs on the quantised EED—needed for Hawking-type results—were only derived recently [25]. What we will do in this paper is to use those QEIs as the basis for a singularity theorem valid in semiclassical quantum gravity. There remains a slight gap. The bounds obtained in [25] hold in flat and curved spacetime backgrounds but in the latter case require a reference state to be chosen. (In Minkowski spacetime we simply adopt the Poincaré invariant vacuum state as the reference.) As we will argue, however, the contribution from the reference state can be neglected on sufficiently small scales, on which the bound is dominated by terms with a universal form (essentially equal to the bound in Minkowski spacetime). This is due to the universal singularity structure of Hadamard states, and the delineation of ‘sufficiently small scales’ depends both on the spacetime and the reference state and may vary over the spacetime. To be slightly more specific, the scales must be small enough that the ‘universal’ terms in the QEI bound outweigh those coming from the renormalised stress energy tensor in the reference state (see the arguments leading up to equation (27)). The upshot is that we assume the bound (2) for solutions to semiclassical Einstein–Klein–Gordon system, with constants Q_m and Q_0 obtained from the QEI and other assumptions, but restricting to functions f whose support extends over sufficiently short intervals of proper time. This timescale is expected to shrink as a singularity is approached. Ideally, it would be derived from the QEI itself—a line of work that will be pursued elsewhere—but in the present paper it must be supplied as an additional piece of information. Thus we will speak of a geodesic being T -regular, where $T : I \rightarrow \mathbb{R}$, if the inequality (2) holds for functions f supported in intervals of width T_t centered at t , allowing t to vary within I (see definition 3.1 for a precise statement).

One approach (which had been our initial plan) is only to assume (2) for f with support duration below a fixed timescale determined by the largest curvature scales encountered along the geodesic γ . This corresponds to T -regularity for a constant function T . In principle a direct application of the results in [15] would now give a singularity theorem predicting geodesic incompleteness with a side assumption that the curvature remain below some threshold chosen sufficiently high that spacetimes violating the assumption might be judged to be ‘singular for practical purposes’. However we found this to require such large extrinsic curvatures at the initial surface as to be of limited practical use. What we are going to do here is to allow the ‘sufficiently small’ timescale to shrink linearly in proper time along the geodesic $\gamma : [0, \tau] \rightarrow \mathcal{M}$, by taking

$$T_t = T_0(1 - t/\tau) \quad (3)$$

for some $T_0 > 0$. On physical grounds, spacetimes for which this is not valid should exhibit very large curvatures along future complete geodesics and therefore also exhibit singular behaviour ‘for practical purposes’. We then employ a partition of unity argument to control the Ricci tensor over the whole of γ , where the bump function in the partition near $t \in [0, \tau]$ has support width less than T_t .

Simplifying slightly, our result is essentially the following. Consider any solution to the semiclassical Einstein–Klein–Gordon equations in which the expected magnitude of the scalar field does not exceed some threshold ϕ_{\max} . It is assumed that the spacetime of the solution is globally hyperbolic and that the state of the quantum field is Hadamard. Let S be a spacelike Cauchy surface in the spacetime and suppose that, for some $T_0, \tau > 0$, all future-directed unit-speed timelike geodesics of length τ emanating normally from S are T -regular where T_t is given by (3). Suppose also that the SEC holds for a period of proper time τ_0 at the start of each such geodesic. If the extrinsic curvature of S is sufficiently large and negative, where the threshold is determined by T_0, τ, τ_0 and the spacetime dimension n , then no future-directed timelike curve emanating from S can have proper duration more than τ and the spacetime is future-timelike geodesically incomplete. Of course, a sufficiently large extrinsic curvature $-K > (n-1)/\tau_0$ at the initial surface will result, by Hawking’s original theorem, in timelike geodesic incompleteness within the timescale τ_0 on which the SEC is assumed. It is important to clarify that there are indeed regimes where our threshold for $-K$ is much lower than $(n-1)/\tau_0$. Another way of interpreting our result is that if the initial extrinsic curvature exceeds our threshold in a solution to the semiclassical Einstein–Klein–Gordon theory obeying the SEC near S , then one or more of the following must occur: (a) the spacetime is timelike geodesically incomplete; (b) the field strength must somewhere exceed ϕ_{\max} ; (c) the timescale on which (2) holds is somewhere smaller than $T_t = T_0(1 - t/\tau)$ at proper time t along a geodesic emanating normally from S . Depending on the choice of parameters, possibilities (b) and (c) may be regarded as indicating the onset of early-universe conditions and therefore of singular behaviour for practical purposes. Of course, it is well known that there are specific models in semi-classical gravity that avoid a cosmological singularity with a bounce—see e.g. [26–29]. This is not in contradiction with our result, but indicates that there are circumstances in which possibilities (b) or (c) come to the fore, and the fact that the initial extrinsic curvature must exceed a given threshold. Apart from cosmological models there are also non-singular semiclassical models of stellar collapse. The singularity is avoided either by avoiding the formation of a trapped surface [30] or the presence of negative energy does not allow the inner horizon to reach the center [31]. It would be interesting to examine the latter models using our methods but such a task is beyond the scope of this paper.

The paper is structured as follows. Section 2 briefly summarises the quantization procedure used in [25] to obtain a QEI on the EED (called a quantum strong energy inequality or QSEI). The QSEI is recalled in section 3, where we also explain why the bounds of the type (2) may be expected on sufficiently small scales, leading to the idea of a T -regular geodesic. In section 4 we define the partition of unity used to discuss averages over long timescales, deferring many details to appendix A. Section 5 presents the singularity theorem that constitutes our main result. It is then shown in section 6 how quantitative information can be extracted concerning the time at which geodesic incompleteness occurs and the required extrinsic curvature at the initial surface. In the context of a toy cosmological model we show that plausible extrinsic curvatures lead to singular behaviour on timescales that exceed the lifetime of our own Universe by as little as one order of magnitude. As a definition of ‘plausible’ parameters, we use values that are predicted to have occurred in our Universe according to Λ CDM cosmology fitted with PLANCK data [32]. We emphasise that the purpose of these calculations is to show that results of the type we have developed do indeed apply beyond the domain of applicability of Hawking’s result, and also as a proof of principle that the quantitative information produced by results of the type we have developed might have practical use. A summary and discussion of our results in section 7 concludes the paper.

We work in n spacetime dimensions unless otherwise stated and employ $[-, -, -]$ conventions in the Misner *et al* classification [33]. That is, the metric signature is $(+, -, -, \dots)$,

the Riemann tensor is defined as $R_{\lambda\eta\nu}{}^{\mu}v^{\nu} = (\nabla_{\lambda}\nabla_{\eta} - \nabla_{\eta}\nabla_{\lambda})v^{\mu}$, and the Einstein equation is $G_{\mu\nu} = -8\pi T_{\mu\nu}$. The d'Alembertian is written $\square_g = g^{\mu\nu}\nabla_{\mu}\nabla_{\nu}$. For the most part, we adopt units in which $G_N = c = 1$.

2. Quantization

Throughout this work, we assume that the spacetime is a smooth n -dimensional Lorentzian manifold (\mathcal{M}, g) that is globally hyperbolic. We consider the minimally coupled classical scalar field ϕ with Lagrangian density

$$L[\phi] = \frac{\sqrt{-g}}{2} ((\nabla\phi)^2 - M^2\phi^2), \quad (4)$$

where M has dimensions of inverse length, equal to the reduced Compton length of massive particles in the corresponding quantum field theory. The field equation and stress–energy tensor are

$$(\square_g + M^2)\phi = 0 \quad (5)$$

and

$$T_{\mu\nu} = (\nabla_{\mu}\phi)(\nabla_{\nu}\phi) + \frac{1}{2}g_{\mu\nu}(M^2\phi^2 - (\nabla\phi)^2), \quad (6)$$

while the EED defined by equation (1) becomes

$$\rho_U = U^{\mu}U^{\nu}(\nabla_{\nu}\phi)(\nabla_{\mu}\phi) - \frac{1}{n-2}M^2\phi^2. \quad (7)$$

Quantization follows exactly the procedure described in reference [25] so we will be brief and direct the reader there for more details. It is based on the algebraic approach (see [34] for a review) and starts by introducing a unital $*$ -algebra $\mathcal{A}(\mathcal{M})$ on our manifold \mathcal{M} , so that self-adjoint elements of $\mathcal{A}(\mathcal{M})$ are observables of the theory. The algebra is generated by elements $\Phi(f)$ where $f \in C_0^{\infty}(\mathcal{M})$, which represent smeared quantum fields and obey the following relations

- **Linearity**

The map $f \rightarrow \Phi(f)$ is complex-linear,

- **Hermiticity**

$$\Phi(f)^* = \Phi(\bar{f}) \quad \forall f \in C_0^{\infty}(\mathcal{M}),$$

- **Field equation**

$$\Phi((\square_g + M^2)f) = 0 \quad \forall f \in C_0^{\infty}(\mathcal{M}),$$

- **Canonical commutation relations**

$$[\Phi(f), \Phi(h)] = iE(f, h)1 \quad \forall f, h \in C_0^{\infty}(\mathcal{M}).$$

Here, $C_0^{\infty}(\mathcal{M})$ denotes the space of smooth functions of compact support from \mathcal{M} to \mathbb{C} and $E(x, y) = E^A(x, y) - E^R(x, y)$ is an antisymmetric bi-distribution equal to the difference of the advanced and retarded Green functions of $\square_g + M^2$, which exist due to global hyperbolicity of the spacetime.

In the algebraic approach, states of the theory are described by linear maps $\omega : \mathcal{A}(\mathcal{M}) \rightarrow \mathbb{C}$ that are normalized ($\omega(1) = 1$) and positive ($\omega(A^*A) \geq 0$ for all $A \in \mathcal{A}(\mathcal{M})$). The expression

$\omega(A)$ is interpreted as the expectation value of $A \in \mathcal{A}(\mathcal{M})$ in state ω . Physical conditions must be specified to narrow down the choice of states, and a particularly well-behaved class of states is given by those that are Hadamard (see [34], and also [35] for a recent improvement to the original definition [36]).

The algebra $\mathcal{A}(\mathcal{M})$ only contains polynomials built from elements of the form $\Phi(f)$. For example, $\mathcal{A}(\mathcal{M})$ contains elements of the form

$$:\Phi^{\otimes 2}:\omega(f \otimes f) = \Phi(f)\Phi(f) - W(f, f)1 \quad (8)$$

for any test function f and any state ω , where

$$W(f, h) = \omega(\Phi(f)\Phi(h)) \quad (9)$$

is the two-point function of ω . These elements have the property that $\langle :\Phi^{\otimes 2}:\omega(f \otimes f) \rangle_\omega = 0$.

To form a Wick square, one needs to replace $f \otimes f \in C_0^\infty(\mathcal{M} \times \mathcal{M})$ by a suitable compactly supported distribution supported on the diagonal (and also address finite renormalisation freedoms). This idea can be implemented in the case where ω is quasifree⁴ and Hadamard and leads to the definition of an extended algebra $\mathcal{W}(\mathcal{M})$, described in [37], that contains $\mathcal{A}(\mathcal{M})$ as a subalgebra, but also contains elements such as Wick polynomials and the smeared stress–energy tensor. For example, the quadratic Wick polynomials in the field and its derivatives needed to define the quantized stress–energy tensor are given as follows.

Let $f^{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}$ be a smooth compactly supported tensor field and define a compactly supported distribution $T^{r,s}[f]$ on $\mathcal{M} \times \mathcal{M}$ by

$$T^{(r,s)}[f](S) = \int_{\mathcal{M}} d\text{Vol} f^{\mu_1 \dots \mu_r \nu_1 \dots \nu_s} \llbracket (\nabla^{(r)} \otimes \nabla^{(s)}) S^{\text{sym}} \rrbracket_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}. \quad (10)$$

Here, $S^{\text{sym}}(x, y) = \frac{1}{2}(S(x, y) + S(y, x))$ is the symmetric part of $S \in C^\infty(\mathcal{M} \times \mathcal{M})$, while $\nabla^{(r)}$ is a symmetrised r th order covariant derivative and the double square brackets $\llbracket \cdot \rrbracket$ in the integrand denote a coincidence limit. Then we obtain a smeared Wick polynomial

$$:\nabla^{(r)} \Phi \nabla^{(s)} \Phi:\omega(f) := :\Phi^{\otimes 2}:\omega(T^{r,s}[f]), \quad (11)$$

which has a vanishing expectation value in the reference state ω . However, as there is no preferred choice of state in a general curved spacetime ([38]; see [39] for discussion) it is desirable to seek prescriptions leading to local and covariant Wick polynomials. This can be done in various ways and the renormalisation freedom cannot be removed completely. Here we follow the axioms of Hollands and Wald [37, 40]. In this scheme, Wick polynomials obey a Leibniz rule for differentiation of products, but the field equation cannot be used freely.

The axioms are enough to define the stress–energy tensor $T_{\mu\nu}$ for perturbatively interacting theories, up to finite renormalisation freedoms, which might be fixed by reference to the expectation values obtained in certain standard states on standard spacetimes, or other criteria. In the case of the linear real scalar field, the finite renormalisation freedom reduces to linear combinations of the tensors $M^4 g_{\mu\nu}$, $M^2 G_{\mu\nu}$, $I_{\mu\nu}$ and $J_{\mu\nu}$ where the latter are conserved symmetric curvature tensors arising by functional derivatives of R^2 and $R_{\mu\nu} R^{\mu\nu}$ with respect to the metric. It is natural to use some of this freedom to choose a prescription in which the Minkowski vacuum state has vanishing expectation value, for example. At any rate, we assume that a prescription has been fixed as part of the physical specification of the matter model. Once this is

⁴ A state is quasifree if all its odd n -point functions vanish and all its even n -point functions can be expanded as sums of products of the two-point function according to Wick's theorem.

done, the semiclassical Einstein equation (SEE) for the Einstein–Klein–Gordon system with vanishing cosmological constant can be formulated as follows.

Definition 2.1. Fix a locally covariant prescription for the stress–energy tensor of the massive minimally coupled scalar field. A *solution to the SEE* is a triple (\mathcal{M}, g, ω) where (\mathcal{M}, g) is a globally hyperbolic spacetime of even dimension $n = 2m$ and ω is a Hadamard state of the quantised minimally coupled Klein–Gordon field on (\mathcal{M}, g) , such that

$$G_{\mu\nu} = -8\pi \langle T_{\mu\nu} \rangle_\omega \quad (12)$$

holds everywhere on \mathcal{M} .

The system (12) is a highly complex set of equations linking the evolution of the metric to that of the two-point function of the quantum state ω . If a prescription has been chosen in which the Minkowski vacuum state has vanishing expected stress–energy tensor, then Minkowski space furnishes a rather trivial example solution. More generally, the analysis of these equations is highly nontrivial. Recently Sanders [41] has succeeded in classifying all static symmetric semiclassical solutions to the Einstein–Klein–Gordon system with spatial topology \mathbb{S}^3 and arbitrary cosmological constant and allowing for nonminimal coupling. Before that, the existence of solutions has been shown in situations with high levels of symmetry, for example in flat FLRW spacetimes (see [42, 43]). In this paper, however, we will formulate a singularity theorem for solutions to the SEE without further discussing the question of existence. Clearly, if (\mathcal{M}, g, ω) is a solution to the SEE, then the Ricci tensor obeys

$$R_{\mu\nu} U^\mu U^\nu = -8\pi \langle \rho_U \rangle_\omega \quad (13)$$

for any timelike unit vector U^μ , where ρ_U is the renormalized EED. As a smeared field, ρ_U takes the form

$$\rho_U(f) = (\nabla_\mu \Phi \nabla_\nu \Phi)(U^\mu U^\nu f) - \Phi^2 \left(\frac{1}{n-2} M^2 f \right), \quad (14)$$

where U is any unit timelike vector field and f is any smooth compactly supported function on \mathcal{M} .⁵

For further manipulations, the following point-splitting formula is useful:

$$\begin{aligned} \langle (\nabla^{(r)} \Phi \nabla^{(s)} \Phi)(f) \rangle_\omega - \langle (\nabla^{(r)} \Phi \nabla^{(s)} \Phi)(f) \rangle_{\omega_0} &= \langle :(\nabla^{(r)} \Phi \nabla^{(s)} \Phi):_{\omega_0}(f) \rangle_\omega \\ &= T^{r,s}[f](W - W_0), \end{aligned} \quad (15)$$

where both ω and ω_0 are quasifree Hadamard states with respective two-point functions W and W_0 . Because ω and ω_0 are Hadamard, the difference $:W: = W - W_0$ is necessarily smooth, so the right-hand side is well-defined. Here, and for the rest of this paper, we denote all quantities normal-ordered relative to ω_0 by $:X:$, rather than $:X:_{\omega_0}$. It follows that

$$\langle \rho_U(f) \rangle_\omega = \langle \rho_U(f) \rangle_{\omega_0} + \langle : \rho_U : (f) \rangle_\omega, \quad (16)$$

where

$$\langle : \rho_U : (f) \rangle_\omega = \int_{\mathcal{M}} d\text{Vol } f \left[\mathbb{I}(\nabla_U \otimes \nabla_U) : W : - \left(\frac{M^2}{n-2} \right) \mathbb{I} : W : \right]. \quad (17)$$

⁵ At nonminimal coupling there are terms such as $\Phi(\square_g + \xi R + M^2)\Phi$. These do not necessarily vanish in the Holands–Wald scheme, but they turn out to be state-independent and therefore cancel in differences of expectation values in Hadamard states.

The quantity appearing in square brackets will be denoted $\langle \rho_U \rangle_\omega$ by a slight abuse of notation.

It is important to note that equation (13) assumes a particular choice to fix finite renormalisation freedoms has been made for ρ_U , inherited from the choices made for the stress–energy tensor. In equation (16) the term ρ_U is independent of such choices because they are manifested in terms that are multiples of the identity operator.

3. Quantum strong energy inequality

We can now turn to the QEIs satisfied by the quantized EED derived in reference [25], which established a variety of lower bounds on averages of $\langle \rho_U \rangle_\omega$ along worldlines or over spacetime volumes, allowing for possibly nonminimal coupling to the scalar curvature. Here, we will employ the worldline result for the minimally coupled massive field.

Let γ be a smooth geodesic parametrized by proper time τ and let U^μ be a timelike unit vector field agreeing with $\dot{\gamma}^\mu$ on γ . For any real valued test function $f \in C_0^\infty(\mathbb{R}, \mathbb{R})$, we write

$$\langle \rho_U \circ \gamma \rangle_\omega(f^2) = \int d\tau f^2(\tau) \langle \rho_U \rangle_\omega(\gamma(\tau)). \quad (18)$$

It was shown in theorem 5 of [25] that the QEI

$$\langle \rho_U \circ \gamma \rangle_\omega(f^2) \geq - \left[\int_0^\infty \frac{d\alpha}{\pi} \phi^*((\nabla_U \otimes \nabla_U)W_0)(\bar{f}_\alpha, f_\alpha) + \frac{M^2}{n-2} \langle \Phi^2 \circ \gamma \rangle_\omega(f^2) \right] \quad (19)$$

holds for all Hadamard states ω , where $f_\alpha(\tau) = e^{i\alpha\tau} f(\tau)$ and $\phi: \mathbb{R} \times \mathbb{R} \rightarrow \mathcal{M} \times \mathcal{M}$ by $\phi(\tau, \tau') = (\gamma(\tau), \gamma(\tau'))$. Consequently the Hadamard-renormalized EED obeys the inequality

$$\begin{aligned} \langle \rho_U \circ \gamma \rangle_\omega(f^2) + \frac{M^2}{n-2} \langle \Phi^2 \circ \gamma \rangle_\omega(f^2) &\geq \langle \rho_U \circ \gamma \rangle_{\omega_0}(f^2) + \frac{M^2}{n-2} \langle \Phi^2 \circ \gamma \rangle_{\omega_0}(f^2) \\ &\quad - \int_0^\infty \frac{d\alpha}{\pi} \phi^*((\nabla_U \otimes \nabla_U)W_0)(\bar{f}_\alpha, f_\alpha) \end{aligned} \quad (20)$$

for all Hadamard states ω , for any choice of Hadamard reference ω_0 with two-point function W_0 , and for all real-valued test functions f .

This bound is not easy to use in practice, due to the lack of closed-form expressions for Hadamard two-point functions that could define a reference state. An alternative would be to derive an ‘absolute’ QSEI that avoids the need for a reference state, and this will be pursued elsewhere. Here, we will instead work around the obstacle by making a physically motivated approximation. Namely, if the support of f is sufficiently small, the integration in the last term of (20) involves the two-point function $W_0(x, y)$ near to coincidence along γ . In this regime, the two-point function is dominated by its leading singularity, which is universal across all Hadamard states and is shared, in particular, by the Minkowski vacuum two-point function of a massless scalar field. For test functions of sufficiently small support—where ‘sufficiently small’ is determined by the reference state ω_0 and the spacetime geometry but *not* the states ω of interest—the bound (20) continues to hold for all ω if the integral on the right-hand side is replaced by

$$\int_0^\infty \frac{d\alpha}{\pi} \int_{\mathbb{R}^2} dt dt' \partial_t \partial_{t'} W_{\text{Mink}}(t, \mathbf{0}; t', \mathbf{0}) \bar{f}_\alpha(t) f_\alpha(t') \quad (21)$$

where

$$W_{\text{Mink}}(x, x') = \hbar \int \frac{d^{n-1}\mathbf{k}}{(2\pi)^{n-1}} \frac{e^{-ik \cdot (x-x')}}{2k} \quad (22)$$

is the massless Minkowski vacuum two-point function and $k^\mu = (\|\mathbf{k}\|, \mathbf{k})$. Planck's constant appears explicitly as it has not been set to unity. The integral may be evaluated as

$$\begin{aligned} & \hbar \int_0^\infty \frac{d\alpha}{\pi} \int \frac{d^{n-1}\mathbf{k}}{(2\pi)^{n-1}} \int_{\mathbb{R}^2} dt dt' \frac{k}{2} e^{-i(\alpha+k)(t-t')} f(t)f(t') \\ &= \hbar \frac{S_{n-2}}{(2\pi)^n} \int_0^\infty d\alpha \int_0^\infty dk k^{n-1} |\hat{f}(\alpha+k)|^2 \\ &= \hbar \frac{S_{n-2}}{(2\pi)^n} \int_0^\infty du \int_0^u dk k^{n-1} |\hat{f}(u)|^2 \\ &= \hbar \frac{S_{n-2}}{(2\pi)^n} \int_0^\infty du \frac{u^n}{n} |\hat{f}'(u)|^2, \end{aligned} \quad (23)$$

where S_{n-2} is the volume of the $(n-2)$ -dimensional standard unit sphere and \hat{f} is the Fourier transform $\hat{f}(u) = \int dt e^{-iut} f(t)$. In the first step, we have used the definition of the transform and integrated out the angular dependence, and in the second we have changed variables from α and k to $u = \alpha + k$ and k . Restricting to spacetimes with even dimension of at least 4, writing $n = 2m$ for $m \in \mathbb{N}$ and $m \geq 2$, we may use the fact that $\hat{f}'(u) = iu\hat{f}(u)$ to obtain

$$\begin{aligned} \hbar \frac{S_{n-2}}{(2\pi)^n} \int_0^\infty du \frac{u^n}{n} |\hat{f}'(u)|^2 &= \hbar \frac{S_{2m-2}}{4m(2\pi)^{2m}} \int_{-\infty}^\infty du |\widehat{f^{(m)}}|^2(u) \\ &= \hbar \frac{\pi S_{2m-2}}{2m(2\pi)^{2m}} \int_{-\infty}^\infty dt |f^{(m)}|^2, \end{aligned} \quad (24)$$

where we have also extended the integral to the full line, using the fact that $|\hat{f}|^2$ is real and even, and used Parseval's theorem in the final step. As we have argued, this expression can be used—at least for sufficiently small support width—as a replacement for the integral in (20). We can say a bit more about the magnitude of this term in relation to the others, because smooth functions f , compactly supported in an interval I of length τ , obey the Rayleigh–Ritz inequality

$$\int_{-\infty}^\infty dt |f^{(m)}|^2 \geq \frac{\lambda_{\min}^{(m)}}{\tau^{2m}} \int_{-\infty}^\infty dt |f|^2, \quad (25)$$

where $\lambda_{\min}^{(m)}$ is the minimum eigenvalue of the operator $(-1)^m d^{2m}/dx^{2m}$ on $(0, 1)$, subject to Dirichlet boundary conditions. (See [44] for other applications of this observation.) By contrast, the first two terms in (20) can be bounded below by

$$\inf_I \left(\langle \rho_U \circ \gamma \rangle_{\omega_0} + \frac{M^2}{n-2} \langle \Phi^2 \circ \gamma \rangle_{\omega_0} \right) \int_{-\infty}^\infty dt |f|^2, \quad (26)$$

so it is clear that the third term dominates when τ is sufficiently small. In other words, the first two terms can be made less in magnitude than any prescribed multiple of the third, for small

enough τ . Thus for any constant $C > 1$, and any $t \in \mathbb{R}$, there is a timescale $T_t > 0$ so that an inequality

$$\langle \rho_U \circ \gamma \rangle_\omega(f^2) + \frac{M^2}{n-2} \langle \Phi^2 \circ \gamma \rangle_\omega(f^2) \geq -C\hbar \frac{\pi S_{2m-2}}{2m(2\pi)^{2m}} \int_{-\infty}^{\infty} dt |f^{(m)}|^2, \quad (27)$$

holds for all Hadamard states ω and all real-valued f supported in the interval $(t - T_t/2, t + T_t/2)$. Thus (27) holds for all real-valued f belonging to

$$\mathcal{D}_T := \{f \in C_0^\infty(\mathbb{R}) : \text{supp } f \subset (t - T_t/2, t + T_t/2) \text{ for some } t \in \mathbb{R}\}. \quad (28)$$

As, in fact, the QEIs derived in reference [25] are not expected to be optimally sharp, we may without loss replace C by 1. The timescale T_t depends on the curve γ , the spacetime geometry, the mass parameter M , and the reference state ω_0 . The effect of the mass parameter may be significant because QEI bounds for massive fields are exponentially smaller than those of massless fields—see [45] for detailed analysis. Therefore our use of the massless bound is quite conservative and may be expected to hold over reasonable length scales. Furthermore, as we are free to choose any Hadamard state ω_0 as the reference we could in principle choose a different reference state for each value of $t \in \mathbb{R}$ so as to maximise T_t .

Of course it is far from trivial to compute T_t in this way and no sufficient conditions on T_t are known for (27) to hold (with $C = 1$) for all real-valued $f \in \mathcal{D}_T$. Such conditions would presumably involve the injectivity radius near $\gamma(t)$ and derivatives of curvature tensors in a neighbourhood thereof. Extrapolating from [46] it may be that only the first three derivatives of the Riemann tensor are relevant at leading order. Lacking sufficient conditions of this type, we turn the problem around and make the following definition.

Definition 3.1. For $\tau \in [0, \infty]$ let $\gamma : [0, \tau] \cap [0, \infty) \rightarrow \mathcal{M}$ be a unit-speed timelike geodesic. Let $T : t \mapsto T_t$ be a strictly positive function on $[0, \tau)$ with dimensions of time. We say that γ is *T-regular* if (a) γ may be extended to a geodesic $[-\tau_*, \tau] \cap \mathbb{R} \rightarrow \mathcal{M}$ where $\tau_* = \sup_{t \in [0, \tau]} (\frac{1}{2}T_t - t)$ and (b) the EED of the real scalar field with mass M obeys

$$\langle \rho_U \circ \gamma \rangle_\omega(f^2) + \frac{M^2}{n-2} \langle \Phi^2 \circ \gamma \rangle_\omega(f^2) \geq -\hbar \frac{\pi S_{2m-2}}{2m(2\pi)^{2m}} \int_{-\infty}^{\infty} dt |f^{(m)}|^2 \quad (29)$$

for all real-valued $f \in \mathcal{D}_T \cap C_0^\infty(-\tau_*, \tau)$ and all Hadamard states ω .

The purpose of condition (a) is to ensure that every interval $(t - T_t/2, t + T_t/2)$ lies in the domain of the geodesic, in order that condition (b) be well-defined (clearly $\tau_* > 0$ because $\tau_0 > 0$). Note that if γ is *T-regular*, then it is also \hat{T} -regular for any strictly positive function $t \mapsto \hat{T}_t \leq T_t$.

The above definition will be important in the singularity theorems for semiclassical quantum gravity to be proved in section 5, where we consider *T-regular* geodesics $\gamma : [0, \tau] \rightarrow \mathcal{M}$ determined by the function $T_t = (1 - t/\tau)T_0$. By definition 3.1, any solution to the SEE satisfies the geometric condition

$$\int dt f^2(t) R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu \leq \hbar \frac{S_{2m-2}}{m(2\pi)^{2m-2}} \int_{-\infty}^{\infty} dt |f^{(m)}|^2 + \frac{4\pi M^2}{m-1} \langle \Phi^2 \circ \gamma \rangle_\omega(f^2) \quad (30)$$

along any *T-regular* unit-speed timelike geodesic, for all $f \in \mathcal{D}_T \cap C_0^\infty(-\tau_*, \tau)$. Note that the timescale T_t on which (29) holds becomes vanishingly small as $t \rightarrow \tau$, which means that the constraint it provides becomes progressively weaker as $\gamma(\tau)$ is approached.

The second term of the lhs of equation (30) is state dependent; however, we can restrict ω to a class of Hadamard states for which the field's magnitude has a finite maximum magnitude.

Definition 3.2. A ϕ_{\max} -solution to the SEE is a solution to the SEE according to definition 2.1 where the state ω is additionally restricted to a class of Hadamard states for which

$$|\langle :\Phi^2 : \circ \gamma \rangle_\omega| \leq \phi_{\max}^2, \quad (31)$$

where ϕ_{\max} is a finite constant.

The above assumptions allow us to make the following statement, which is immediate from equation (30) using definition 3.2.

Proposition 3.3. For any ϕ_{\max} -solution to the SEE according to definition 3.2 on a T -regular geodesic according to definition 3.1 the inequality

$$\int dt f^2(t) R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu \leq \frac{\hbar S_{2m-2}}{m(2\pi)^{2m-2}} \|f^{(m)}\|^2 + \frac{4\pi M^2 \phi_{\max}^2}{m-1} \|f\|^2 \quad (32)$$

holds for all real-valued $f \in \mathcal{D}_T \cap C_0^\infty(-\tau_*, \tau)$, where the norms are those of $L^2(0, \infty)$.⁶

4. Partition of unity

Let $\tau \in (0, \infty]$ and suppose $\gamma: [0, \tau] \cap [0, \infty) \rightarrow \mathcal{M}$ is a T -regular unit-speed timelike geodesic. According to proposition 3.3 a ϕ_{\max} -solution to the SEE satisfies the QEI of equation (32) on γ . By definition, this means that the QEI holds for test functions of sufficiently small support, where a neighborhood of $t \in [0, \tau)$ is deemed sufficiently small if it is contained in $(t - \frac{1}{2}T_t, t + \frac{1}{2}T_t) \cap [-\tau_*, \tau)$. However, the singularity theorems require control over averages of the EED taken over the whole of $[0, \tau)$. These averages may be addressed by choosing a partition of unity to break the long-term average into pieces each of which is sufficiently small that (32) holds.

Accordingly, we seek smooth compactly supported functions $\phi_n \in C_0^\infty(-\tau_*, \tau)$ such that

$$\sum_{n=1}^{\infty} \phi_n^2 = 1 \quad (33)$$

holds on $[0, \tau)$ and with $\text{supp } \phi_n \subset [t_{n-1}, t_n]$, where the strictly increasing sequence t_n ($n \in \mathbb{N}_0$) is chosen so that $t_n \rightarrow \tau$ as $n \rightarrow \infty$ and

$$[t_{n-1}, t_n] \subset [t_n - T_{t_n}/2, t_n + T_{t_n}/2] \quad (34)$$

for each n , which implies $\phi_n \in \mathcal{D}_T \cap C_0^\infty(-\tau_*, \tau)$.

For each $f \in C_0^\infty(0, \tau)$, we may make a decomposition $f^2 = \sum_{n=1}^{\infty} (f\phi_n)^2$ and apply proposition 3.3 to each function $f\phi_n$ separately. Summing up, we obtain the inequality

$$\int_0^\tau dt f^2(t) R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu \leq Q_0 \|f\|^2 + Q_m \sum_{n=1}^{\infty} \|(f\phi_n)^{(m)}\|^2, \quad (35)$$

for any $f \in C_0^\infty(0, \tau)$, where the sum contains only finitely many nonzero terms and

⁶ We only use the inequality for real-valued f , but it extends immediately to complex-valued f on considering the real and imaginary parts separately and replacing f^2 by $|f(t)|^2$ on the lhs.

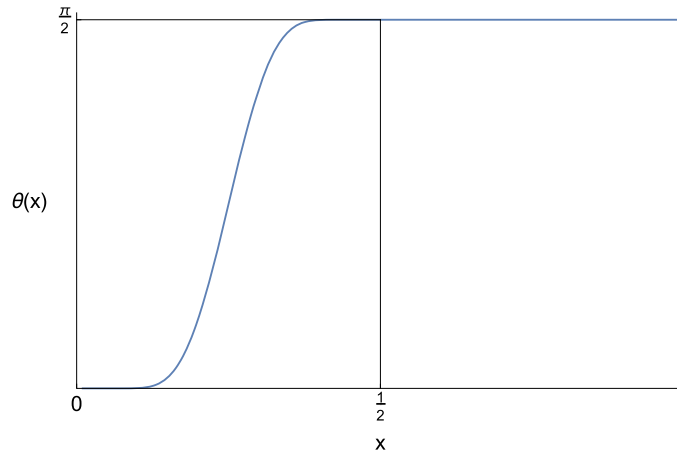


Figure 1. The function $\theta(x)$ defined in equation (38).

$$Q_m = \frac{\hbar S_{2m-2}}{(2\pi)^{2m-2}}, \quad \text{and} \quad Q_0 = \frac{4\pi M^2 \phi_{\max}^2}{m-1} \quad (36)$$

(recall that $m \geq 1$ so there is no conflict in the above definition). For suitably chosen ϕ_n , the second term in (35) may be estimated in terms of a weighted Sobolev norm.

As the construction of the functions ϕ_n and the derivation of the required estimates is quite complex, we summarise the main points here and provide the details in appendix A. First, in appendix A.1, we will construct functions ϕ_n obeying the conditions above, for any positive, nonincreasing continuous function T_t . Here we choose t_n to satisfy the equation $t_{n+1} - t_n = \frac{1}{2}T_{t_{n+1}}$ for $n \geq 0$ and $t_0 = -\frac{1}{2}T_0$, from which $t_1 = 0$ follows; it is shown in particular that $t_n \rightarrow \tau$ as $n \rightarrow \infty$. The corresponding ϕ_n are defined by

$$\phi_n(t) = \begin{cases} F((t - t_n)/T_{t_n} + 1/2) & t < t_n, \\ G((t - t_n)/T_{t_{n+1}}) & t \geq t_n, \end{cases} \quad (37)$$

where F and G are the functions $F(x) = \sin(\theta(x))$ and $G(x) = \cos(\theta(x))$, and $\theta \in C^\infty(\mathbb{R})$ is the smooth nonnegative function given by

$$\theta(x) = A \int_0^x H(x')H(1/2 - x')e^{-1/x'} e^{-1/(1/2-x')} dx', \quad (38)$$

where $H(x)$ is the Heaviside step function. Here, $A \approx 3.2392 \times 10^4$ is chosen so that $\theta(1/2) = \pi/2$. The function θ interpolates smoothly from 0 to $\pi/2$, as shown in figure 1.

In appendix A.2 we will consider examples of T_t for both finite and infinite intervals. It will also be shown that other examples can be obtained by reverse engineering the function T_t from a sequence t_n . A function of interest, appropriate for a finite interval $[0, \tau]$, is $t \mapsto T_t = T_0(1 - t/\tau)$. For that choice we have

$$t_n = \tau \left(1 - \left(\frac{2\tau}{T_0 + 2\tau} \right)^{n-1} \right), \quad (39)$$

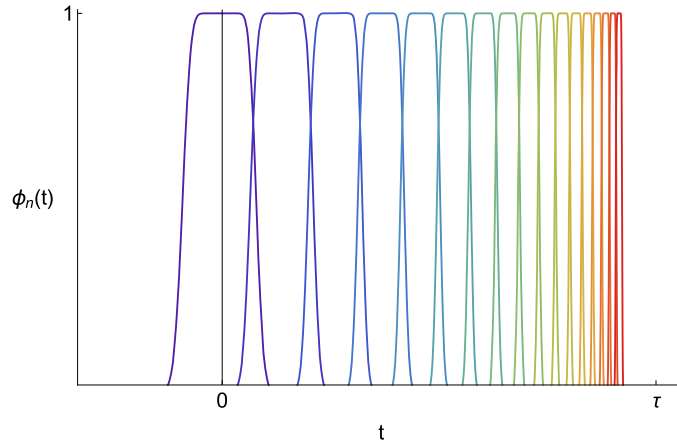


Figure 2. The series of functions ϕ_n for the linear $T_t = T_0(1 - t/\tau)$. The center of each function is given by t_n (equation (39)) and their support shrinks to a point as $t \rightarrow \tau$.

for all $n \in \mathbb{N}_0$. The functions ϕ_n have vanishing support width in the limit $t \rightarrow \tau$ as illustrated in figure 2.

In appendix A.3 we estimate the second term of equation (35). Our aim is to estimate this quantity in terms of (weighted) Sobolev norms of f , keeping reasonably explicit control of any constants appearing. We will do that both for a partition of unity corresponding to a general T -regular geodesic and a specific choice of T_t for a finite interval. In particular, for the choice of $T_t = T_0(1 - t/\tau)$ and choosing τ_0 with $0 < \tau_0 < t_2 = \tau T_0/(T_0 + 2\tau)$, we will show that

$$\begin{aligned} \sum_{n=1}^{\infty} \|(f\phi_n)^{(m)}\|^2 &\leq 2 \left[\sum_{j=0}^m H_j \binom{m}{j} \left(\frac{\|\chi f^{(m-j)}\|}{(2t_2)^j} + c^j \left\| \frac{(1-\chi)f^{(m-j)}}{(\tau - \cdot)^j} \right\| \right) \right]^2 \\ &\equiv |||f|||^2, \end{aligned} \quad (40)$$

where χ is the characteristic function of $[0, \tau_0]$,

$$H_i = \max\{\|F^{(i)}\|_{\infty}, \|G^{(i)}\|_{\infty}\}, \quad (41)$$

and

$$c = \frac{2\tau^2}{(2\tau - T_0)T_0}. \quad (42)$$

The upshot of this analysis is expressed by the following theorem.

Theorem 4.1. *Let $0 < \tau < \infty$, $T_0 > 0$ and $0 < \tau_0 < \tau T_0/(T_0 + 2\tau)$, define $T_t = T_0(1 - t/\tau)$. For any ϕ_{\max} -solution to the SEE according to definition 3.2 on a T -regular geodesic according to definition 3.1, the inequality*

$$\int d\tau f^2(t) R_{\mu\nu} \dot{\gamma}^{\mu} \dot{\gamma}^{\nu} \leq Q_m |||f|||^2 + Q_0 \|f\|^2 \quad (43)$$

holds for all real-valued $f \in W_0^m(0, \tau)$, where $|||f|||^2$ is given in equation (40) and the norms are those of $L^2(0, \infty)$.

Here, we recall that the Sobolev space $W_0^m(0, \tau)$ is the completion of the space $C_0^\infty(0, \tau)$ in the norm $\|f\|_m = (\|f\|^2 + \tau^{2m}\|f^{(m)}\|^2)^{1/2}$. In particular, its elements are functions with $m - 1$ absolutely continuous⁷ derivatives which all vanish (including the function itself) at the endpoints, and whose m th distributional derivative may be identified with a element of $L^2(0, \tau)$.

Proof. For $f \in C_0^\infty(0, \tau)$, the inequality (43) follows immediately from proposition 3.3 using equations (35) and (40). Now let $f \in W_0^m(0, \tau)$ and let $f_n \in C_0^\infty(0, \tau)$ be a sequence converging to f in $W_0^m(0, \tau)$. We may apply (43) to each f_n , and observe that $\int d\tau f_n(\tau)^2 R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu \rightarrow \int d\tau f(\tau)^2 R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu$ and $\|f_n\| \rightarrow \|f\|$ as $n \rightarrow \infty$. Therefore (43) will apply to f provided one also has $\|f_n\| \rightarrow \|f\| < \infty$. As $\|\chi f_n^{(m-j)}\| \rightarrow \|\chi f^{(m-j)}\|$ for all $0 \leq j \leq m$, attention can be focused on the remaining terms contributing to $\|\cdot\|$. These are controlled using the higher-order Hardy inequality (proved in appendix A.4)

$$\int_{\tau_0}^{\tau} \frac{|h(t)|^2}{(\tau - t)^{2j}} dt \leq \frac{4^j}{((2j - 1)!!)^2} \int_{\tau_0}^{\tau} |h^{(j)}(t)|^2 dt, \quad (44)$$

which holds for all absolutely continuous functions h whose first $j - 1$ derivatives are also absolutely continuous, and which obey $h(\tau) = h'(\tau) = \dots = h^{(j-1)}(\tau) = 0$ and hence for all $h \in W_0^j(0, \tau)$. It is then elementary to see that $|\|f_n\| - \|f\|| \leq \|f_n - f\| \leq \text{const} \times \|f_n - f\|_m \rightarrow 0$, completing the proof. \square

5. The singularity theorem

We now apply methods from [15] to prove a singularity theorem applicable to ϕ_{\max} -solutions (\mathcal{M}, g, ω) to the SEE. We will assume that there is a Cauchy surface S in \mathcal{M} from which all normally emanating future-directed timelike geodesics of length τ are T -regular, with $T_t = T_0(1 - t/\tau)$ for T_0 positive. Furthermore, we assume that the SEC holds for a short period of proper time along these geodesics after S (this corresponds to ‘scenario 1’ in [15]). We will prove that if the initial extrinsic curvature scalar on S is sufficiently negative then every such geodesic contains a focal point. It follows by standard arguments (summarised in [15]) that the spacetime is future timelike geodesically incomplete—indeed, no future-directed timelike curve leaving S can have length greater than τ . In full detail, we will prove the following singularity theorem, which closely follows the structure of theorems 4.2 and 4.4 of reference [15].

Theorem 5.1. *Let (\mathcal{M}, g, ω) be a ϕ_{\max} -solution of the SEE according to definition 3.2 and let S be a smooth spacelike Cauchy surface in \mathcal{M} . Suppose that:*

- (a) *There exist constants $\tau > 0$ and $T_0 > 0$ so that every unit-speed timelike geodesic $\gamma : [0, \tau] \rightarrow \mathcal{M}$ emanating normally from S is T -regular, with $T_t = T_0(1 - t/\tau)$ (see definition 3.1)*
- (b) *There exist constants $\rho_0 \leq 0$ and $\tau_0 \in (0, \tau T_0/(T_0 + 2\tau))$ so that, along each such γ , the inequality $R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu|_{\gamma(t)} \leq \rho_0$ holds for $t \in [0, \tau_0]$;*
- (c) *The extrinsic curvature K of S satisfies*

⁷ The absolutely continuous functions on $[0, \tau]$ those arising as integrals of L^1 (and hence in particular L^2) functions.

$$-K \geq \min \left\{ \frac{n-1}{\tau_0}, \nu_*(T_0, \tau_0, \tau, \rho_0) \right\}, \quad (45)$$

with $\nu_*(T_0, \tau_0, \tau, \rho_0)$ given by equation (54) below.

Then no future-directed timelike curve emanating from S has length greater than τ and \mathcal{M} is future timelike geodesically incomplete.

The idea of the proof is to show that each unit-speed geodesic $\gamma : [0, \tau] \rightarrow \mathcal{M}$ emanating normally from S contains a focal point. Under the hypotheses of the theorem, let $\rho(t) = R_{\mu\nu}U^\mu U^\nu|_{\gamma(t)}$ for some such γ ,⁸ which defines a smooth function on $[0, \tau]$ because the state ω is Hadamard. By proposition 2.2 of reference [15], which brings together results from [12, 16] and references therein, γ contains a focal point to S if there is a piecewise smooth $g : [0, \tau] \rightarrow \mathbb{R}$ with $g(0) = 1$ and $g(\tau) = 0$, such that

$$J[g] \leq -K|_{\gamma(0)}, \quad (46)$$

where

$$J[g] = \int_0^\tau ((n-1)\dot{g}(t)^2 + g(t)^2\rho(t)) dt. \quad (47)$$

We proceed to estimate $J[g]$ for the specific piecewise smooth function $g \in C^{m-1}([0, \tau])$ given by

$$g(t) = \begin{cases} 1 & t \in [0, \tau_0) \\ I(m, m; (\tau - t)/(\tau - \tau_0)) & t \in [\tau_0, \tau], \end{cases} \quad (48)$$

where $I(p, q; x)$ [commonly written $I_x(p, q)$] is the regularised incomplete Beta function⁹. Note that $g(0) = 1$, $g(\tau) = 0$. A plot of g for $m = 2$ is given in figure 3, which also shows the related function $\varphi : [0, \tau] \rightarrow \mathbb{R}$ defined by

$$\varphi(t) = \begin{cases} I(m, m; t/\tau_0) & t \in [0, \tau_0) \\ 1 & t \in [\tau_0, \tau]. \end{cases} \quad (49)$$

The advantage of these functions for our purposes is that their L^2 -norms and those of their first derivatives are known in closed form [15, 47]. To be specific, the L^2 -norms on the interval $[0, \tau]$ are

$$\|\chi_{[\tau_0, \tau]}g\|^2 = (\tau - \tau_0)A_m, \quad \|g'\|^2 = \frac{B_m}{\tau - \tau_0}, \quad (50)$$

(see the proof of lemma 4.1 in [15]) where $\chi_{[a, b]}$ is the characteristic function of $[a, b]$ and the constants A_m and B_m are

$$A_m = \frac{1}{2} - \frac{(2m)!^4}{4(4m)!m!^4}, \quad B_m = \frac{(2m-2)!^2(2m-1)!^2}{(4m-3)!(m-1)!^4} \quad (51)$$

⁸ By the SEE, $\rho(t) = -8\pi\langle\rho_U\rangle_\omega$, which is a slightly unfortunate collision of notation between the present paper and [15].

⁹ The function g was also used in section 4.1 of reference [15], where it was denoted f .

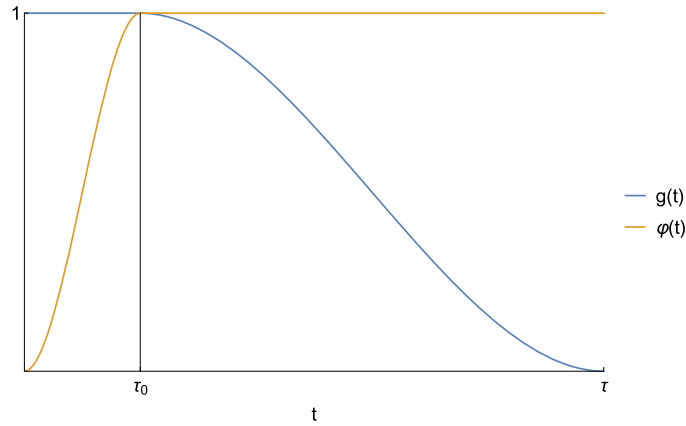


Figure 3. Plot of the functions $g(t)$ and $\varphi(t)$ defined in equations (48) and (49) respectively.

Table 1. The first few values of the constants A_m and B_m .

m	1	2	3	4
A_m	1/3	13/35	181/462	521/1287
B_m	1	6/5	10/7	700/429

for $m \in \mathbb{N}$. The first few relevant values are tabulated in table 1. For the function φ we have $\|\chi_{[0,\tau_0]}\varphi\|^2 = \tau_0 A_m$. It is then easily seen that for the product $f = g\varphi$ we have

$$\|f\|^2 = A_m \tau. \quad (52)$$

As the function $f = g\varphi$ belongs to $W_0^m(0, \tau)$, and γ satisfies the hypotheses of theorem 4.1 due to assumption (a) of theorem 5.1, it may be shown that

$$\int_0^\tau \rho(t) f(t)^2 dt \leq Q_0 A_m \tau + Q_m H_m(T_0, \tau_0, \tau), \quad (53)$$

where $H_m(T_0, \tau_0, \tau) = \|f\|^2$ is given by equation (99) and the details of the derivation are discussed in appendix A.5. Now it is straightforward to prove the following lemma.

Lemma 5.2. *Under the hypotheses of theorem 5.1, if $\gamma : [0, \tau] \rightarrow \mathcal{M}$ is a unit-speed geodesic emanating normally from S then the functional J defined by (47) satisfies the estimate*

$$\begin{aligned} J[g] &\leq \nu_*(T_0, \tau_0, \tau, \rho_0) := Q_m H_m(T_0, \tau_0, \tau) + Q_0 A_m \tau \\ &\quad + (n-1) \frac{B_m}{\tau - \tau_0} + \rho_0 \tau_0 (1 - A_m) \end{aligned} \quad (54)$$

for the function g given by (48). Consequently, if $-K|_{\gamma(0)} \geq \nu_*$ then γ contains a focal point to S .

Proof. Using $g^2 = (\varphi g)^2 + (1 - \varphi^2)g^2 = f^2 + (1 - \varphi^2)$ and equation (53) we have

$$\begin{aligned}
\int_0^\tau dt g(t)^2 \rho(t) &= \int_0^\tau dt \rho(t) f(t)^2 + \int_0^\tau dt (1 - \varphi^2) \rho(t) \\
&\leq Q_m H_m(T_0, \tau_0, \tau) + Q_0 A_m \tau + \int_0^{\tau_0} dt (1 - \varphi^2) \rho(t) \\
&\leq Q_m H_m(T_0, \tau_0, \tau) + Q_0 A_m \tau + \rho_0 \tau_0 (1 - A_m),
\end{aligned} \tag{55}$$

where we have also used the estimates for the Sobolev norms of g , φ and f and assumption (b) of theorem 5.1. Using the estimate of $\|\dot{g}\|^2$ from equation (50) the proof of (54) is complete and the remaining statement follows by [15, proposition 2.2] as discussed above. \square

The proof of our main result is now straightforward.

Proof of theorem 5.1. Using assumptions (a)–(c) and lemma 5.2, all unit-speed timelike geodesics of length τ emanating normally from S contain a focal point to S . It follows that every unit-speed geodesic emanating normally from S that does not contain focal points has length strictly less than τ . Theorem 5.1 now follows by standard reasoning, as summarised e.g. in proposition 2.4 (b) of [15]. \square

6. A quantitative application using cosmological parameters

To conclude, we will show that the theorems given above can be used quantitatively. We will use parameters drawn from the cosmology of our Universe and show that the time-reversed version of our theorem predicts a past singularity, or singularity-like behaviour, on reasonable cosmological timescales. It should be borne in mind that this prediction relates to our SEE coupled to a single scalar quantum field. Thus the aim is not to make predictions about our Universe but rather to show that our singularity theorem can make physically plausible quantitative statements about a toy cosmology. This was also our approach in [13, 15] to which we will refer for some details.

The relevant information about our own Universe is drawn from the Λ CDM cosmology model fitted with data from the PLANCK experiment [32]. In particular, the ratios of dust and dark energy densities to the critical density are taken respectively as $\Omega_m = 0.31$ and $\Omega_\Lambda = 0.69$ at the present time. According to the model fitted with these parameters, the SEC was last satisfied when the Universe's age was approximately 2.41×10^{17} s (at redshift parameter $z_* = 0.642$), after which dark energy became dominant. The Hubble parameter at that time was $H_* = 3.14 \times 10^{-18} \text{ s}^{-1}$ (for details on this calculation see [15]). Also at that time, the four-dimensional Ricci scalar had magnitude $|R_*| = 5.7 \times 10^{-35} \text{ s}^2$. What we draw from this is that (this model of) our own Universe contains a Cauchy surface with extrinsic curvature $K = H_*$ towards the future, and that the SEC was obeyed for some period beforehand.

Now let us examine the quantitative predictions of the time-reversed version of theorem 5.1, in the four-dimensional case $m = 2$. The idea is the Cauchy surface S in the theorem should correspond to the $z_* = 0.642$ hypersurface in the Λ CDM model. Several parameters are required: M (the mass of the field, expressed as the inverse reduced Compton length), ϕ_{\max} (the maximum magnitude of the scalar field), timescales T_0 and τ that determine the function T_t , together with parameters ρ_0 and τ_0 relating to the period where SEC is assumed. Recall that our assumption is that the SEC should hold in the form $R_{\mu\nu}\dot{\gamma}^\mu\dot{\gamma}^\nu \leq \rho_0 \leq 0$ for at least a period of time τ_0 just before the surface S . We set $\rho_0 = 0$, so our assumption becomes simply the statement that SEC held for at least a period τ_0 . For M , we use values corresponding to

three elementary particles: the pion, proton¹⁰ and Higgs. In each case ϕ_{\max} is determined using the square root of the Wick square in a Minkowski spacetime KMS state of a temperature that is 1% of the reduced Compton temperature $T_{\text{Compton}} = c\hbar M/k$ of the particle concerned, where k is Boltzmann's constant. The temperature T_{Compton} defines a scale beyond which the model cannot be trusted and, following [13, 15] provides a heuristic basis for the value

$$\phi_{\max}^2 \sim 10^{-2} \frac{c^4}{G_N} (M\ell_{\text{Pl}})^2 K_1(100), \quad (56)$$

where K_ν is a modified Bessel function of the second kind and ℓ_{Pl} is the Planck length. We should note that equation (56) is sensitive to the choice of temperature $10^{-2}T_{\text{Compton}}$ and that ϕ_{\max} would change significantly if this is raised or reduced. For the moment, we leave the timescales T_0 , τ and τ_0 as free parameters. Let us recall that T_t is the timescale over which the Minkowski QEI provides a valid estimate on the EED in the spacetime, and in particular that T_0 sets that timescale at the surface S . For theorem 5.1 we must assume $\tau_0 < t_2 = \tau T_0/(T_0 + 2\tau)$.

The time-reversed version of theorem 5.1 asserts that, provided the past-directed unit-speed geodesics emanating normally from S are T -regular, with $T_t = T_0(1 - t/\tau)$, past geodesic incompleteness occurs within timescale τ if the extrinsic curvature of S (to the future) exceeds $\min\{3/\tau_0, \nu_*(T_0, \tau_0, \tau, \rho_0)\}$, where ν_* was defined in equation (54). As $\rho_0 = 0$,

$$\nu_*(T_0, \tau_0, \tau, 0) = Q_2 H_2(T_0, \tau_0, \tau) + Q_0 A_2 \tau + \frac{3B_2}{\tau - \tau_0}, \quad (57)$$

where $A_2 = 13/35$, $B_2 = 6/5$ and, restoring dimensions, the constants Q_2 and Q_0 are given by

$$Q_2 = \hbar \frac{G_N}{2\pi c^5} = 4.60 \times 10^{-88} \text{ s}^2, \quad Q_0 = \frac{4\pi G_N M^2 \phi_{\max}^2}{c^2}. \quad (58)$$

The values of Q_0 for the three elementary particles based on the ϕ_{\max} calculation described are given in table 2, while the function $H_2(T_0, \tau_0, \tau)$ is evaluated in appendix A.5 as

$$\begin{aligned} H_2(T_0, \tau_0, \tau) = 2 \left[\sqrt{\frac{12}{\tau_0^3}} + \sqrt{\frac{12}{(\tau - \tau_0)^3}} + 2H_1 \left(\sqrt{\frac{3}{10\tau_0 t_2^2}} + c \sqrt{\frac{12}{(\tau - \tau_0)^3}} \right) \right. \\ \left. + H_2 \left(\sqrt{\frac{13\tau_0}{560t_2^4}} + c^2 \sqrt{\frac{13}{3(\tau - \tau_0)^3}} \right) \right]^2, \end{aligned} \quad (59)$$

where the numerical coefficients are $H_1 = 10.87$ and $H_2 = 234.6$ and $c = 2\tau^2/((2\tau - T_0)T_0)$

The required extrinsic curvature depends on the three remaining independent parameters T_0 , τ_0 and τ . We will consider situations in which τ , the timescale within which the singularity occurs, is much greater than T_0 , the timescale on which the Minkowski QEI is valid in the curved spacetime near S . This implies that $\tau_0 < t_2 \approx T_0/2 \ll \tau$ and $c \approx \tau/T_0 \gg 1$. If the extrinsic curvature at S exceeds $3/\tau_0$, then Hawking's theorem already implies that past timelike geodesic incompleteness occurs within time τ_0 of S , i.e. in the region where SEC

¹⁰ The proton is of special interest in cosmology as the majority of baryonic matter is in the form of hydrogen. Of course the proton, as a spin-1/2 particle, is not described by a scalar field and we emphasize again that our aim here is to obtain reasonable order of magnitude parameters rather than to accurately model cosmology. Nonetheless, the QEIs for fermionic fields that have been developed (e.g. [48, 49]) take a similar form to those applying to scalar fields. It seems reasonable to expect that the same would be true for a QSEI, although no such bound has yet been derived for fermionic fields.

Table 2. The temperature scale and Q_0 for three elementary particles with different masses, the pion, the Higgs and the proton.

Particle	Mass in kg	T_{Compton} in K	Q_0 in s^{-2}	τ in s	ν_* in s^{-1}	$\min T_0$ in s
Pion	2×10^{-28}	1.56×10^{12}	2.38×10^{-40}	2.02×10^{20}	3.57×10^{-20}	1.05×10^{-10}
Proton	1.67×10^{-27}	1.089×10^{13}	5.60×10^{-37}	4.16×10^{18}	1.73×10^{-18}	1.51×10^{11}
Higgs	2.23×10^{-25}	1.45×10^{15}	1.78×10^{-28}	2.33×10^{14}	3.09×10^{-14}	1.14×10^{-13}

holds. This rather trivial result is avoided by restricting to the situation where $\nu_* \ll 3/\tau_0$. Our theorem now predicts that past timelike geodesic completeness fails within time τ provided $K > \nu_*$.

For convenience, let us assume that the first term of equation (57) is much smaller than the other two (as actually happens for a broad range of parameters). Then the required extrinsic curvature is given by a combination of the last two terms,

$$\nu_*(T_0, \tau_0, \tau, 0) \approx Q_0 A_2 \tau + \frac{3B_2}{\tau} \quad (60)$$

different for each particle, but roughly independent of T_0 and τ_0 provided we remain in the regime mentioned above. The value of τ that minimizes ν_* and the corresponding value of ν_* are

$$\tau = \sqrt{3B_2/(Q_0 A_2)} = 3.11 Q_0^{-1/2}, \quad \nu_* = \sqrt{12A_2 B_2 Q_0} = 2.31 Q_0^{1/2}. \quad (61)$$

Evaluating these expressions for the pion gives $\tau = 2.02 \times 10^{20}$ s and $\nu_* = 3.57 \times 10^{-20} \text{ s}^{-1}$, while for the proton we have $\tau = 4.16 \times 10^{18}$ s and $\nu_* = 1.73 \times 10^{-18} \text{ s}^{-1}$. Both values of ν_* are smaller than the value $H_* = 3.14 \times 10^{-18} \text{ s}^{-1}$ drawn from Λ CDM cosmology. Therefore, in either of these models, we could conclude the presence of a past singularity within timescale τ if the extrinsic curvature took the value H_* on S . Note that the timescales are larger, by 1 order of magnitude (proton) or 3 orders of magnitude (pion) than the age of our Universe at $z_* = 0.642$. Particularly for the proton, these results show that the theorem provides physically plausible quantitative estimates. In the Higgs case, we have $\tau = 2.33 \times 10^{14}$ s and $\nu_* = 3.09 \times 10^{-14} \text{ s}^{-1}$, very much larger than H_* . Similar behaviour was observed in an earlier hybrid model [15].

The parameters τ_0 and T_0 are still free, though must be chosen to be much less than τ and to ensure that all the subsidiary assumptions made above still hold. Let us examine the condition that the first term in equation (57) be much smaller than the others, which sum to the approximate value of ν_* given in equation (61) in our situation of interest. Taking $\tau_0 \approx T_0/2$ the assumption $T_0 \ll \tau$ implies that $Q_2 H_2(T_0, \tau_0, \tau) \approx 4.77 \times 10^5 Q_2 \tau T_0^{-4}$. Using the above values, the first term of (57) is at least two orders of magnitude smaller than the sum of the other two provided $T_0 \geq 1.31 \times 10^{-20} Q_0^{-1/4}$, giving values of 1.06×10^{-10} s for the pion, 1.51×10^{-11} s for the proton and 1.14×10^{-13} s for the Higgs. The largest of these values corresponds to a length scale of a few centimeters. By comparison, in the Λ CDM model at the $z_* = 0.642$ surface, the four-dimensional Ricci scalar $|R_*| = 5.7 \times 10^{-35} \text{ s}^2$ gives a crude indication that curvature scales are on the order of $|R_*|^{-1/2} = 1.3 \times 10^{17}$ s, while the Hubble parameter sets a timescale $H_*^{-1} = 3.2 \times 10^{17}$ s of a similar order. As the minimum values of T_0 are many orders of magnitude below those scales it seems very reasonable to accept the validity of the Minkowski QEI. The τ , ν_* and T_0 values for each particle are summarized in table 2.

If, instead of taking $\tau_0 \approx T_0/2$, we allow τ_0 to be much less than T_0 , then the contribution to H_2 equal to $24Q_2/\tau_0^3$ might become dominant. Requiring this term to remain at least two orders of magnitude smaller than $4.77 \times 10^5 Q_2 \tau T_0^{-4}$ we find a minimum value of τ_0 for which our approximations remain valid. This reasoning gives a minimum value of τ_0 for the pion 1.46×10^{-21} s, for the proton 4.00×10^{-22} s and for the Higgs 1.53×10^{-23} s. In other words, the SEC need be assumed to hold only for a fleeting instant (this does not exclude it holding for a much longer period!). Furthermore, in all cases considered we have $3/\tau_0 \gtrsim 6/T_0 \gg 6/\tau = 0.84\nu_*$ so we are well outside the regime (by several orders of magnitude) in which Hawking's result would predict geodesic incompleteness.

Finally, we note that there is still a lot of freedom that could be exploited. For example, if we multiply the value of τ from that given in equation (61) by a factor of a , the value of ν_* is multiplied by a factor of $a/2 + 1/(2a)$. In the case of the proton we may use this freedom to lower the singularity timescale for a proton to about 5.3 times the age of our Universe at the $z_* = 0.642$ surface, while the required extrinsic curvature remains less than H_* .

7. Conclusions

In this work we presented the first proof of a semiclassical singularity theorem based directly on a QEI satisfied by a quantum field. In previous works we have derived, on the one hand, a singularity theorem with a weakened energy condition [15] and, on the other, a quantum strong energy inequality (QSEI) [25]. Here we have connected those pieces, showing how a suitable energy condition for a singularity theorem can be derived from such a QEI. The main challenge we overcame was the need to approximate the curved spacetime QEI by its Minkowski counterpart, which necessitated a partition of unity argument into pieces of vanishingly short duration towards the end of the geodesic segments under consideration, thus allowing for a possible blow-up in the curvature scales.

In particular, we showed that a solution to the SEE with a massive scalar field cannot be future timelike complete if four conditions hold: (i) an assumption essentially requiring that the local temperature of the scalar field remains below a specified threshold, (ii) all the unit speed timelike geodesics of length τ emanating normally from a spacelike hypersurface S are T -regular for a function $T_t = T_0(1 - t/\tau)$ that decreases linearly to the future (i.e. the Minkowski QEI provides a reliable estimate of the curved spacetime QEI on timescales given by T_t at time t), (iii) the SEC holds for a short period just after S and (iv) the extrinsic curvature of S should obey a specific inequality. Failure of any of conditions (a) and (b) might be regarded as indicating singular behaviour for many practical purposes, provided the temperature threshold is high enough—for instance, failure of (ii) indicates that the curvature scales shrink to zero faster than linearly along at least one geodesic leaving S .

To gain some insight into the potential utility of our theorem, it was applied using data drawn from Λ CDM cosmology and the PLANCK experiment [32], with the mass of the scalar field taken equal to that of the pion, the proton and the Higgs particle in turn. For the proton, we found physically reasonable parameter values so that an extrinsic curvature of a magnitude found at redshift $z = 0.642$ in the Λ CDM model would guarantee a singularity within an order of magnitude of the actual age of the Universe at that time. The SEC need only be assumed for a fleeting instant—far shorter than would be needed for Hawking's result to predict a singularity. It should be emphasised that our result holds across *all* solutions to the SEE satisfying the assumptions, so it is not at all surprising that the timescale to singularity is longer than in the physical cosmology. Nonetheless, it is gratifying that the simple model produces such physically plausible results.

A drawback in our result is the assumption of T -regularity, which works around the fact that curved spacetime QEI in its current form requires the specification of a reference state. In fact, we believe it likely that our approach ends up overestimating the threshold on the extrinsic curvature needed to guarantee geodesic incompleteness. A more satisfactory approach would require a suitable ‘absolute’ version of the QEI from [25] that does not require a reference state. Even then it would be necessary to estimate various terms in the Hadamard parametrix. Such a computation has been done in the case of the quantum weak energy inequality [46] based on an absolute QEI from [50]. The QSEI presents additional challenges as the bound is state dependent even in the minimally coupled scalar field case and is the topic of ongoing work.

The singularity theorem presented here concerns timelike geodesic incompleteness. In reference [15] we derived singularity theorems with weakened energy conditions for both timelike and null geodesics. However, QEIs along null geodesics in four-dimensional spacetime do not exist, as was shown by an explicit counterexample in [51]. The counterexample can be circumvented if a momentum cut-off is imposed, and it was recently shown in reference [52] that a conjectured null QEI with a UV momentum cutoff [53] is sufficient to prove a null geodesic singularity theorem using methods from [15]. However, the QEI involved remains to be proven and any curvature corrections estimated.

Finally, it is important to note that the SEE is a highly nontrivial system and relatively little is known about its solution space except where a high level of symmetry is imposed. In the case of static spacetimes, let us note that all of the solutions studied in [41] are timelike geodesically complete (in all cases the metric is simply $g = dt^2 - a^2 h$ where h is the round three-sphere metric and $a > 0$ is constant). One might wonder whether this indicates a contradiction with our theorems, but this is not the case: because the solutions in [41] also obey the timelike convergence condition, there can be no spacelike Cauchy surface whose extrinsic curvature is bounded away from zero—because otherwise the Hawking singularity theorem would imply the failure of timelike geodesic completeness. Therefore our theorem does not apply to these solutions and there is accordingly no contradiction. Solutions have also been derived in the case of cosmological spacetimes—see e.g. [42, 43, 54]. It would be of interest to examine the behavior of solutions obtained in those works and the range of initial conditions that predict a primordial singularity.

Acknowledgments

The authors would like to thank Rafael Sorkin for a useful conversation regarding the use of the semiclassical Einstein equation, and Daniel Siemssen for a conversation about the relationship between our results and those of [41]. This work is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 744037 ‘QuEST’. E-AK is supported by the ERC Consolidator Grant QUANTIVIOL.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Appendix A. Details concerning the partition of unity

In this appendix we provide the details of the partition of unit construction summarized in section 4.

A.1. Construction of the partition of unity

We aim to construct functions ϕ_n obeying equations (33) and (34), starting from a given function $t \mapsto T_t$. A first task is to show that a sequence t_n of the required type can be found.

Lemma A.1. *Suppose that $t \mapsto T_t$ is strictly positive, nonincreasing and continuous on $[0, \tau)$ where $0 < \tau \leq \infty$, with $T_t \rightarrow 0$ as $t \rightarrow \tau$ if $\tau < \infty$. Then there is a strictly increasing sequence t_n ($n \in \mathbb{N}_0$) in $[0, \tau)$ satisfying (34), with $t_0 = -\frac{1}{2}T_0$, $t_1 = 0$, $t_{n+1} - t_n = \frac{1}{2}T_{t_{n+1}}$ for $n \geq 0$ and $t_n \rightarrow \tau$ as $n \rightarrow \infty$. Furthermore, defining $n : [0, \tau) \rightarrow \mathbb{N}$ so that $t' \in [t_{n(t')}, t_{n(t')+1})$ for all $t' \in [0, \tau)$, the following estimates hold for all $t \in [0, \tau)$:*

$$t_{n(t)+1} \leq t + \frac{1}{2}T_t, \quad T_{t_{n(t)+1}} \geq T_{t+T_t/2}. \quad (62)$$

Proof. We first show that there is a sequence $t_n \in [0, \tau)$ obeying $t_1 = 0$ and $t_{n+1} - t_n = \frac{1}{2}T_{t_{n+1}}$ for $n \geq 1$ (clearly this holds also for $n = 0$ if we set $t_0 = -\frac{1}{2}T_0$). If $\tau < \infty$ then set $T_t = 0$ for $t \geq \tau$, thus extending T_t as a continuous function. Suppose inductively that the first n terms of the sequence can be constructed for $n \geq 1$ and all lie in $[0, \tau)$, noting that this holds for $n = 1$. The function $h(t) = t - \frac{1}{2}T_t$ is continuous on $[0, \infty)$, strictly increasing, obeys $h(t_n) < t_n$ and $h(t) \geq t - \frac{1}{2}T_{t_n}$ for $t \geq t_n$. Thus $h(\tau) > t_n$ and $h(t_n + \frac{1}{2}T_{t_n}) \geq t_n$. By the intermediate value theorem, the equation $h(t) = t_n$ has a solution $t \in (t_n, t_n + \frac{1}{2}T_{t_n}] \cap (0, \tau)$, which is in fact the unique solution in (t_n, ∞) . Therefore t_n is well-defined as a strictly increasing sequence t_n in $[0, \tau)$.

As T_t is nonincreasing, the sequence t_n also has the property $t_{n+1} - t_n = \frac{1}{2}T_{t_{n+1}} \leq \frac{1}{2}T_{t_n} = t_n - t_{n-1}$. It follows that $[t_{n-1}, t_{n+1}] \subset [t_n - T_{t_n}/2, t_n + T_{t_n}/2]$.

If $\tau < \infty$, the sequence is increasing and bounded above, so $t_n \rightarrow t_* = \sup_n t_n \leq \tau$, whereupon $T_{t_n} \rightarrow 0$ as $n \rightarrow \infty$ and hence $T_{t_*} = 0$ by continuity, showing that $t_* \geq \tau$. Thus $t_n \rightarrow \tau$. Similarly, if $\tau = \infty$ but the sequence t_n is bounded, then $t_n \rightarrow t_* < \infty$ and we again deduce $T_{t_*} = 0$, contradicting the assumption that T_t is strictly positive on $(0, \tau)$.

To establish (62) one notes that $t_{n(t)} \leq t < t_{n(t)+1}$ by definition, so

$$t_{n(t)+1} \leq t + t_{n(t)+1} - t_{n(t)} = t + \frac{1}{2}T_{t_{n(t)+1}} \leq t + \frac{1}{2}T_t, \quad (63)$$

as T_t is non-increasing, from which we also deduce $T_{t_{n(t)+1}} \geq T_{t+T_t/2}$. \square

We remark that one also has the very simple estimate $t_{n(t)+1} \leq t + t_2$ for all t , on noting that $t_{n(t)+1} \leq t + \frac{1}{2}T_{t_2} = t + t_2$ for $t \geq t_2$, and trivially, that $t_{n(t)+1} = t_2 \leq t + t_2$ for $t < t_2$. The functions ϕ_n ($n \in \mathbb{N}$) corresponding to the sequence t_n are defined by equation (37), where $F(x) = \sin(\theta(x))$ and $G(x) = \cos(\theta(x))$, with θ given by equation (38). We must check that they are indeed smooth functions satisfying the conditions (33) and (34).

Clearly, each ϕ_n takes values in $[0, 1]$ with $\phi_n(t_n) = 1$. Additionally, ϕ_n is clearly smooth everywhere except perhaps $t = t_n$; but as all derivatives of $\theta(x)$ vanish as $x \rightarrow 0$ or $x \rightarrow \frac{1}{2}$, it is easily seen that each ϕ_n is smooth at t_n with all its derivatives vanishing there.

As $\phi_n(t)$ vanishes identically for $t < t_n - T_{t_n}/2 = t_{n-1}$ and for $t > t_n + T_{t_{n+1}}/2 = t_{n+1}$, we have $\text{supp } \phi_n \subset [t_{n-1}, t_{n+1}]$. Thus, on the interval $[t_n, t_{n+1}]$ only ϕ_n and ϕ_{n+1} can be nonzero, and one has

$$\phi_n(t)^2 + \phi_{n+1}(t)^2 = \cos^2[\theta((t - t_n)/T_{t_{n+1}})] + \sin^2[\theta((t - t_{n+1})/T_{t_{n+1}} + \frac{1}{2})] = 1$$

because $t_{n+1}/T_{t_{n+1}} - 1/2 = t_n/T_{t_{n+1}}$. Summarising, we have constructed a sequence ϕ_n ($n \geq 1$) of smooth compactly supported functions taking values in $[0, 1]$, with $\text{supp } \phi_n \subset [t_{n-1}, t_{n+1}]$ and satisfying (33) on $\cup_{n=1}^{\infty} [t_n, t_{n+1}] = [0, \tau)$.

A.2. Specific examples of the sequence t_n

Finite interval. For a first example, define $T_t = T_0(1 - t/\tau)$ on $[0, \tau)$ for $0 < \tau < \infty$. The equation $t_{n+1} - t_n = \frac{1}{2}T_t$ can be rearranged to

$$t_{n+1} = \frac{t_n + T_0/2}{1 + T_0/(2\tau)}, \quad (64)$$

which is solved uniquely, subject to $t_1 = 0$, by

$$t_n = \tau \left(1 - \left(\frac{2\tau}{T_0 + 2\tau} \right)^{n-1} \right) \quad (65)$$

for all $n \in \mathbb{N}_0$. In particular, $t_2 = \tau T_0/(T_0 + 2\tau)$ and $t_0 = -\frac{1}{2}T_0$. Defining $n : [0, \tau) \rightarrow \mathbb{N}$ so that $t' \in [t_{n(t')}, t_{n(t')+1})$, one finds

$$n(t') = 1 + \left\lfloor \frac{\log(1 - t'/\tau)}{\log(2\tau/(T_0 + 2\tau))} \right\rfloor, \quad (66)$$

from which $t_{n(t)+1}$ may be calculated. Here $\lfloor x \rfloor$ is the largest integer no greater than x . For convenience, one may use the estimate (62) to give

$$t_{n(t)+1} \leq \left(1 - \frac{T_0}{2\tau} \right) t + \frac{T_0}{2} \in [T_0/2, \tau) \quad (67)$$

and therefore

$$T_{t_{n(t)+1}} \geq \frac{(\tau - t)(2\tau - T_0)T_0}{2\tau^2} = \frac{\tau - t}{c}, \quad c = \frac{2\tau^2}{(2\tau - T_0)T_0} \quad (68)$$

holds for all $t \in [0, \tau)$. For $0 \leq t < t_2$, one has $n(t) = 1$ and therefore $T_{t_{n(t)+1}} = T_{t_2} = 2t_2$ holds for $t \in [0, t_2)$.

Infinite interval. As an example where $\tau = \infty$, consider $T_t = T_0 e^{-\lambda t}$ for some constant $\lambda > 0$. Then the equation

$$t_n - t_{n-1} = \frac{1}{2}T_{t_n} \quad (69)$$

may be written as a recurrence relation in terms of Lambert's W -function [55, section 4.13] by

$$t_{n+1} = t_n + \lambda^{-1} W \left(\frac{1}{2} \lambda T_0 e^{-\lambda t_n} \right), \quad t_1 = 0. \quad (70)$$

In particular, $t_2 = W(\frac{1}{2}\lambda T_0)$ obeys $1 < e^{\lambda t_2} < 1 + \frac{1}{2}\lambda T_0$.

To obtain asymptotic approximations to t_n , it is convenient to write $t_n = s_n/\lambda$ where the sequence s_n obeys

$$s_{n+1} - s_n = \mu e^{-s_{n+1}}, \quad s_1 = 0 \quad (71)$$

and $\mu = \frac{1}{2}\lambda T_0$. Observe that the sequence $S_n = \log(\mu n + e^{-\mu})$ satisfies

$$S_{n+1} - S_n = \mu \int_n^{n+1} \frac{dr}{\mu r + e^{-\mu}} > \frac{\mu}{\mu(n+1) + e^{-\mu}} = \mu e^{-s_{n+1}}, \quad (72)$$

$$S_1 = \log(\mu + e^{-\mu}) > 0.$$

Thus $S_1 > s_1$ and, because

$$S_{n+1} - s_{n+1} > S_n - s_n + \mu (e^{-S_{n+1}} - e^{-s_{n+1}}), \quad (73)$$

a simple induction argument shows that $S_n > s_n$ for all integer $n \geq 1$, i.e. one has

$$s_n < \log(\mu n + e^{-\mu}) \quad n \geq 1. \quad (74)$$

Now the relation (71) implies

$$s_n = \mu \sum_{r=2}^n e^{-s_r} \quad (75)$$

so the upper bound on s_n gives

$$\begin{aligned} s_n &> \sum_{r=2}^n \frac{\mu}{\mu n + e^{-\mu}} = \psi(n+1 + \mu^{-1} e^{-\mu}) - \psi(2 + \mu^{-1} e^{-\mu}) \\ &= \log(n) - \psi(2 + \mu^{-1} e^{-\mu}) + O(1/n) \end{aligned} \quad (76)$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

Thus $s_n = \log(\mu n) + O(1)$ as $n \rightarrow \infty$ and the upper and lower bounds on s_n , together with (71), give

$$\begin{aligned} \frac{\mu}{\mu n + e^{-\mu}} &< s_{n+1} - s_n < \frac{\mu \exp(\psi(2 + \mu^{-1} e^{-\mu}))}{\exp(\psi(n+2 + \mu^{-1} e^{-\mu}))} \\ &= \frac{\mu \exp(\psi(2 + \mu^{-1} e^{-\mu}))}{\mu(n+1/2) + e^{-\mu}} (1 + O(n^{-2})) \end{aligned} \quad (77)$$

i.e. $s_{n+1} - s_n = O(1/n)$. That is, our original sequence t_n satisfies

$$t_n = \lambda^{-1} \log\left(\frac{1}{2} \lambda T_0 n\right) + O(1), \quad t_{n+1} - t_n = O(1/n) \quad (78)$$

as $n \rightarrow \infty$.

Reverse engineering. To conclude this section, we note that other example sequences can be obtained by reverse engineering the function T_t from a required sequence t_n . Suppose a strictly increasing sequence t_n ($n \in \mathbb{N}_0$) is given, with $t_1 = 0$ and $t_n \rightarrow \tau$ as $n \rightarrow \infty$. Choose any homeomorphism $k : [0, \infty) \rightarrow [0, 2\tau/T_0 + 1)$ that obeys $k(0) = 0$ and $k(1) = 1$, and provides an interpolation of the sequence t_n according to

$$t_n = \frac{1}{2}(k(n) - 1)T_0. \quad (79)$$

Then it is easily seen, noting that $k^{-1}(2t_n/T_0 + 1) = n$, that this sequence corresponds to

$$T_t = 2t + T_0 - T_0 k(k^{-1}(1 + 2t/T_0) - 1) \quad (80)$$

which, if k is differentiable, obeys $T'_0 = 2(1 - k'(0)/k'(1))$.

For instance, the sequence (65) arises in this way via the homeomorphism $k : [0, \infty) \rightarrow [0, 2\tau/T_0 + 1)$ given by

$$k(x) = 1 + \frac{2\tau}{T_0} \left(1 - \left(\frac{2\tau}{T_0 + 2\tau} \right)^{x-1} \right), \quad (81)$$

while the sequence

$$t_n = \frac{T_0}{2}(\psi(n+1) + \gamma - 1) = \begin{cases} \frac{1}{2}T_0 \left(-1 + \sum_{r=1}^n r^{-1} \right) & n \geq 1 \\ -\frac{1}{2}T_0 & n = 0 \end{cases} \quad (82)$$

where γ is the Euler–Mascheroni constant, arises (via the homeomorphism of $[0, \infty)$ given by $k(x) = \psi(x+1) + \gamma$) from

$$T_t = 2t + T_0 (1 - \gamma - \psi(\psi^{-1}(1 - \gamma + 2t/T_0) - 1)). \quad (83)$$

A.3. Weighted Sobolev inequalities

Returning to a general partition of unity to start with, we can proceed to estimate the sum appearing on the lhs of equation (35).

Exchanging the integral with the sum we can write, for any $f \in C_0^\infty(0, \tau)$,

$$\begin{aligned} \sum_{n=1}^{\infty} \|(f\phi_n)^{(m)}\|^2 &= \int_0^\tau \sum_{n=1}^{\infty} \left| \sum_{i=0}^m \binom{m}{i} f^{(m-i)} \phi_n^{(i)} \right|^2 dt \\ &= \int_0^\tau \sum_{i=0}^m \sum_{j=0}^m \binom{m}{i} \binom{m}{j} \overline{h_i f^{(m-i)} h_j f^{(m-j)}} P_{ij} dt, \\ P_{ij}(t) &= \sum_{r=1}^{\infty} \frac{\phi_r^{(i)}(t) \phi_r^{(j)}(t)}{h_i(t) h_j(t)}, \end{aligned} \quad (84)$$

where the nonnegative functions $h_j(t)$, to be chosen below, have dimensions of $[\text{time}]^{-j}$ to make the components P_{ij} dimensionless. For each $t \geq 0$, there are at most two nonzero terms in the sum defining P_{ij} , namely the terms $r = n(t)$ and $r = n(t) + 1$, where $n(t')$ is defined so that $t' \in [t_{n(t')}, t_{n(t')+1})$. (For example, if t_n is related to an interpolating function k as in equation (79), then

$$n(t) = \lfloor k^{-1}(2t/\tau_c + 1) \rfloor, \quad (85)$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x .) Therefore the sum defining P_{ij} is finite. Using the elementary inequality that $x^\dagger P x \leq \|P\|_{\max} (\sum_i |x_i|)^2$ where $\|P\|_{\max} = \max_{i,j} |P_{ij}|$, the integral in (84) can be bounded as

$$\sum_{n=1}^{\infty} \|(f\phi_n)^{(m)}\|^2 \leq \int_0^\tau \left(\sum_{i=0}^m \binom{m}{i} h_i(t) |f^{(m-i)}(t)| \right)^2 \|P(t)\|_{\max} dt. \quad (86)$$

There is considerable freedom to choose the functions h_i . A convenient choice is to set

$$h_i(t) = \frac{H_i}{T_{t_{n(t)+1}}}, \quad H_i = \max\{\|F^{(i)}\|_\infty, \|G^{(i)}\|_\infty\}. \quad (87)$$

To see why, note that

$$\phi_n^{(i)}(t) = \begin{cases} s_n^i F^{(i)}((t - t_n)/T_{t_n} + 1/2) & t < t_n, \\ s_{n+1}^i G^{(i)}((t - t_n)/T_{t_{n+1}}) & t \geq t_n, \end{cases} \quad (88)$$

where $s_n = 1/T_{t_n}$. At each $t \geq 0$, $\phi_{n(t)}^{(i)}(t) = s_{n(t)+1}^i G((t - t_n)/T_{t_{n+1}})$, while $\phi_{n(t)+1}^{(i)}(t) = s_{n(t)+1}^i F((t - t_{n+1})/T_{t_{n+1}})$. Thus, taking $h_i(t) = T_0^{-i} s_{n(t)+1}^i = T_{t_{n(t)+1}}^{-i}$, one has $|P_{ij}(t)| \leq 2$ and hence the estimates

$$\begin{aligned} \sum_{n=1}^{\infty} \|(f\phi_n)^{(m)}\|^2 &\leq 2 \left\| \sum_{j=0}^m \frac{H_j}{T_{t_{n(t)+1}}^j} \binom{m}{j} |f^{(m-j)}| \right\|^2 \\ &\leq 2 \left[\sum_{j=0}^m H_j \binom{m}{j} \left\| \frac{f^{(m-j)}}{T_{t_{n(t)+1}}^j} \right\| \right]^2 \end{aligned} \quad (89)$$

for all $f \in C_0^\infty(0, \tau)$, where we have used the triangle inequality in the last step.

To proceed, let us consider the specific partition of unity induced by $T_t = T_0(1 - t/\tau)$ on the interval $[0, \tau)$, now assuming $\tau < \infty$. Suppose that $0 < \tau_0 < t_2$ and let χ be the characteristic function of $[0, \tau_0]$. Writing $f^{(m-j)} = \chi f^{(m-j)} + (1 - \chi)f^{(m-j)}$ and again using the triangle inequality,

$$\left\| \frac{f^{(m-j)}}{T_{t_{n(t)+1}}^j} \right\| \leq \left\| \frac{\chi f^{(m-j)}}{T_{t_{n(t)+1}}^j} \right\| + \left\| \frac{(1 - \chi)f^{(m-j)}}{T_{t_{n(t)+1}}^j} \right\|, \quad (90)$$

also using the fact that $T_{t_{n(t)+1}} \geq 2t_2$ for $t \in [0, \tau_0]$ and otherwise using the estimate $T_{t_{n(t)+1}} \geq (\tau - t)/c$, (see (68) and (68)) one has

$$\sum_{n=1}^{\infty} \|(f\phi_n)^{(m)}\|^2 \leq 2 \left[\sum_{j=0}^m H_j \binom{m}{j} \left(\frac{\|\chi f^{(m-j)}\|}{(2t_2)^j} + c^j \left\| \frac{(1 - \chi)f^{(m-j)}}{(\tau - \cdot)^j} \right\| \right) \right]^2 \quad (91)$$

for all $f \in C_0^\infty(0, \tau)$. This is precisely the inequality (40) needed to complete the proof of theorem 4.1.

A.4. A higher order Hardy inequality

Let U and V be measurable real-valued functions on $[0, \infty)$ and suppose that $B > 0$ defined by

$$B^2 = \sup_{x>0} \left(\int_x^\infty |U(t)|^2 dt \right) \left(\int_0^x \frac{dt}{|V(t)|^2} \right)$$

is finite. Where necessary, it is understood that the convention $0 \cdot \infty = 0$ is in force. Then there exists $C > 0$ such that the weighted Hardy inequality

$$\int_0^\infty \left| U(x) \int_0^x f(t) dt \right|^2 dx \leq C^2 \int_0^\infty |V(x)f(x)|^2 dx. \quad (92)$$

holds for measurable, locally integrable f and the least value of the constant C for which this is true obeys $B \leq C \leq 2B$. In particular, the finiteness of the right-hand side implies that the left-hand side is finite. See [56] for an elementary proof. Considerably more general results are known.

We apply this inequality in the case $U(x) = x^{-n}H(1 - x)$, $V(x) = x^{-(n-1)}H(1 - x)$ where H is the Heaviside function, for which

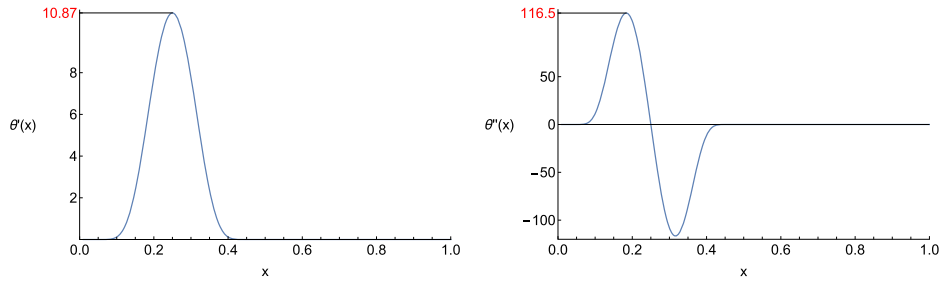


Figure 4. Plots of the first (left) and second (right) derivatives of function θ (see equation (38)). Their maximum values are indicated in red.

$$B^2 = \sup_{x>0} \frac{H(1-x)(x^{-(2n-1)} - 1)}{(2n-1)} \frac{x^{2n-1}}{(2n-1)} = \frac{1}{(2n-1)^2}$$

and therefore

$$\int_0^1 \left| \frac{1}{x^n} \int_0^x f(t) dt \right|^2 dx \leq \frac{4}{(2n-1)^2} \int_0^1 \frac{|f(x)|^2}{x^{2(n-1)}} dx. \quad (93)$$

As the antiderivatives of absolutely integrable functions are precisely the absolutely continuous functions, inequality (93) can be rephrased as the statement that for every $h \in AC([0, 1])$ obeying $h(0) = 0$, one has the inequality

$$\left\| x \mapsto \frac{h(x)}{x^n} \right\|^2 \leq \frac{4}{(2n-1)^2} \left\| x \mapsto \frac{h'(x)}{x^{2(n-1)}} \right\|^2, \quad (94)$$

where h plays the role of $h(x) = \int_0^x f(t) dt$ in (93). Iterating, one finds

$$\left\| x \mapsto \frac{h(x)}{x^n} \right\|^2 \leq \frac{4^n}{((2n-1)!!)^2} \|h^{(n)}\|^2 \quad (95)$$

for all $h \in AC([0, 1])$ such that each $h^{(j)} \in AC([0, 1])$ with $h^{(j)}(0) = 0$ for $0 \leq j \leq n-1$. Inequality (44) is now immediate by an elementary change of integration variable.

See e.g. [57] for more general (though less explicit) higher order Hardy inequalities.

A.5. Computation of $H_m(T_0, \tau_0, \tau)$

In section 5 we need to compute $H_m(T_0, \tau_0, \tau) = \|f\|^2$, for the function f built from incomplete beta functions

$$f(t) = \begin{cases} I(m, m; t/\tau_0) & t \in [0, \tau_0) \\ I(m, m; (\tau - t)/(\tau - \tau_0)) & t \in [\tau_0, \tau]. \end{cases} \quad (96)$$

Elementary manipulations give

Table 3. The values of the constants $A_{m,j}$ and $C_{m,j}$ for $m = 2$.

j	0	1	2
$A_{2,j}$	13/35	6/5	12
$C_{2,j}$	12	12	13/3

$$\|\chi f^{(m-j)}\|^2 = \frac{A_{m,m-j}}{\tau_0^{2(m-j)-1}}, \quad A_{m,k} := \int_0^1 I^{(k)}(m, m; x)^2 dx \quad (97)$$

and

$$\left\| \frac{(1-\chi)f^{(m-j)}}{(\tau-\cdot)^j} \right\|^2 = \frac{C_{m,j}}{(\tau-\tau_0)^{2m-1}}, \quad C_{m,j} = \int_0^1 x^{-2j} I^{(m-j)}(m, m; x)^2 dx. \quad (98)$$

Here, conventions have been chosen so that $A_{m,0} = A_m$ and $C_{m,0} = C_m$, in the notation used in [15] and section 5. Substituting these results in (40) gives

$$H_m(T_0, \tau_0, \tau) := 2 \left[\sum_{j=0}^m H_j \binom{m}{j} \left(\sqrt{\frac{A_{m,m-j}}{\tau_0^{2(m-j)-1} (2t_2)^{2j}}} + c^j \sqrt{\frac{C_{m,j}}{(\tau-\tau_0)^{2m-1}}} \right) \right]^2 \quad (99)$$

for the function f in (96). Here, the dependence on T_0 enters via c and t_2 .

To make this bound quantitative, one needs to evaluate the constants $A_{m,j}$ and $C_{m,j}$. Our main interest will be in the case $m = 2$, for which the relevant values are tabulated in table 3. Using these values together with equation (99) gives

$$H_2(T_0, \tau_0, \tau) = 2 \left[H_0 \left(\sqrt{\frac{12}{\tau_0^3}} + \sqrt{\frac{12}{(\tau-\tau_0)^3}} \right) + 2H_1 \left(\sqrt{\frac{3}{10\tau_0 t_2^2}} + c \sqrt{\frac{12}{(\tau-\tau_0)^3}} \right) + H_2 \left(\sqrt{\frac{13\tau_0}{560t_2^4}} + c^2 \sqrt{\frac{13}{3(\tau-\tau_0)^3}} \right) \right]^2. \quad (100)$$

From equation (41) we have $H_0 = 1$ and

$$H_1 = \sup \theta'(x), \quad H_2 = \sup \theta'(x)^2 + \sup \theta''(x). \quad (101)$$

Now $\theta'(x) = e^{-1/x} e^{-1/(1/2-x)}$ on $[0, 1/2]$ and vanishes otherwise. Differentiating, $\theta''(x) = \theta'(x)(1-4x)/(1-2x)^2$, so it is easily seen that the maximum of θ' occurs at $x = 1/4$, giving $H_1 = 10.87$ (see figure 4). Differentiating further,

$$\theta'''(x) = \theta'(x) \frac{48x^4 - 48x^3 + 32x^2 - 10x + 1}{x^4(-1+2x)^4} \quad (102)$$

and one may determine that the maximum of θ'' occurs at $x = \frac{1}{4} - \frac{\sqrt{6\sqrt{13}-21}}{12}$. Consequently $H_2 = 234.6$.

More generally, closed forms for $A_{m,0} = A_0$ and $C_{m,0} = C_0$ are known for all $m \in \mathbb{N}$. For general $k \geq 1$, we have

$$\begin{aligned}
A_{m,k} &= \frac{1}{B(m,m)^2} \int_0^1 \left(\frac{d^{k-1}}{dx^{k-1}} x^{m-1} (1-x)^{m-1} \right)^2 dx \\
&= \frac{\Gamma(2m)^2}{\Gamma(m)^2(m-k)!^2} \int_0^1 \left(\sum_{r=0}^{m-1} \frac{(1-m)_r(m)_r}{(m-k+1)_r} \frac{x^{m-k+r}}{r!} \right)^2 dx \\
&= \frac{\Gamma(2m)^2}{\Gamma(m)^2(m-k)!^2(2(m-k)+1)} \\
&\quad \times \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{(1-m)_r(m)_r}{(m-k+1)_r} \frac{(1-m)_s(m)_s}{(m-k+1)_s} \frac{(2(m-k)+1)_{r+s}}{(2(m-k+1))_{r+s}} \frac{1}{r!s!} \\
&= \frac{\Gamma(2m)^2}{\Gamma(m)^2(m-k)!^2(2(m-k)+1)} \\
&\quad \times F_{1:1:1}^{1:2:2} \left(\begin{matrix} 2(m-k)+1 : m, 1-m; m, 1-m \\ 2(m-k+1) : m-k+1; m-k+1 \end{matrix} ; 1, 1 \right), \tag{103}
\end{aligned}$$

(note that the series actually terminate after m terms) where the general Kampé de Fériet function [58, 59] in 2 variables is given by

$$F_{B:D;G}^{A:C;F} \left(\begin{matrix} (a) : (c); (f) \\ (b) : (d); (g) \end{matrix} ; x, y \right) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{((a))_{m+n}((c))_m((f))_n}{((b))_{m+n}((d))_m((g))_n} \frac{x^m y^n}{m!n!}, \tag{104}$$

in which (a) , (c) etc are sequences of length A , C etc, and the notation $((c))_m = \prod_{i=1}^C (c_i)_m$ denotes the product of Pochhammer symbols. A similar calculation yields

$$C_{m,j} = \frac{\Gamma(2m)^2}{j!^2 \Gamma(m)^2} F_{1:1:1}^{1:2:2} \left(\begin{matrix} 1 : m, 1-m; m, 1-m \\ 2 : j+1; j+1 \end{matrix} ; 1, 1 \right). \tag{105}$$

In certain cases, these expressions can be evaluated in terms of more familiar special functions. For example, direct calculation gives

$$\int_0^1 x^{\lambda-1} I'(\mu, \mu'; x)^2 dx = \frac{1}{B(\mu, \mu')^2} \int_0^1 x^{\lambda-1} x^{2\mu-2} (1-x)^{2\mu'-2} dx \tag{106}$$

$$= \frac{B(\lambda + 2\mu - 2, 2\mu' - 1)}{B(\mu, \mu')^2} \tag{107}$$

so with $\mu' = \mu$ and $\lambda = 3 - 2\mu$,

$$\int_0^1 x^{-2(\mu-1)} I'(\mu, \mu'; x)^2 dx = \frac{B(1, 2\mu - 1)}{B(\mu, \mu)^2} = \frac{1}{(2\mu - 1)B(\mu, \mu)^2}, \tag{108}$$

giving $C_{m,m-1}$ in the case $\mu = m$.

The coefficients $C_{m,m}$ can also be obtained. In the notation of [47], one has

$$\int_0^1 x^{\lambda-1} I(\mu, \mu', x)^2 dx = B(\lambda, 1) \mathcal{B}(\lambda, 1, \mu, \mu', \mu, \mu') \tag{109}$$

where

$$\mathcal{B}(\lambda, \lambda', \mu, \mu', \nu, \nu') := \frac{1}{B(\lambda, \lambda')} \int_0^1 t^{\lambda-1} (1-t)^{\lambda'-1} I(\mu, \mu'; t) I(\nu, \nu'; t) dt, \tag{110}$$

and one may read off from that reference a formula for the right-hand side in terms of a generalized hypergeometric function:

$$\int_0^1 x^{\lambda-1} I(\mu, \mu', x)^2 dx = \frac{1}{\lambda} \left(1 - \frac{2B(\lambda + 2\mu, \mu')}{\mu B(\mu, \mu')^2} {}_3F_2 \left(\begin{matrix} 1 - \mu', \mu, \lambda + 2\mu \\ 1 + \mu, \lambda + 2\mu + \mu' \end{matrix}; 1 \right) \right), \quad (111)$$

which was proved for nonnegative parameters λ , μ and μ' . Setting $\mu' = \mu$ and analytically continuing the above formula to $\lambda = 1 - 2\mu$ one finds

$$\int_0^1 x^{-2\mu} I(\mu, \mu, x)^2 dx = \frac{1}{2\mu - 1} \left(\frac{2}{\mu^2 B(\mu, \mu)^2} {}_3F_2 \left(\begin{matrix} 1, \mu, 1 - \mu \\ 1 + \mu, 1 + \mu \end{matrix}; 1 \right) - 1 \right), \quad (112)$$

yielding a closed form for $C_{m,m}$ on setting $\mu = m$.

ORCID iDs

Christopher J Fewster  <https://orcid.org/0000-0001-8915-5321>

Eleni-Alexandra Kontou  <https://orcid.org/0000-0003-4409-8188>

References

- [1] Penrose R 1965 Gravitational collapse and space-time singularities *Phys. Rev. Lett.* **14** 57–9
- [2] Hawking S W 1966 The occurrence of singularities in cosmology *Proc. R. Soc. A* **294** 511–21
- [3] Pirani F A E 1956 On the physical significance of the Riemann tensor *Acta Phys. Polon.* **15** 389–405
Pirani F A E 2009 Republication of: on the physical significance of the Riemann tensor *Gen. Relativ. Gravit.* **41** 1215
- [4] Epstein H, Glaser V and Jaffe A 1965 Nonpositivity of the energy density in quantized field theories *Nuovo Cimento* **36** 1016
- [5] Curiel E 2017 A primer on energy conditions *Einstein Stud.* **13** 43–104
- [6] Kontou E-A and Sanders K 2020 Energy conditions in general relativity and quantum field theory *Class. Quantum Grav.* **37** 193001
- [7] Senovilla J M M 2018 Singularity theorems and their consequences (arXiv:1801.04912 [gr-qc])
Senovilla J M M 1998 Singularity theorems and their consequences *Gen. Relativ. Gravit.* **30** 701
- [8] Tipler F J 1978 Energy conditions and spacetime singularities *Phys. Rev. D* **17** 2521–8
- [9] Borde A 1987 Geodesic focusing, energy conditions and singularities *Class. Quantum Grav.* **4** 343–56
- [10] Roman T A 1988 On the ‘averaged weak energy condition’ and Penrose’s singularity theorem *Phys. Rev. D* **37** 546–8
- [11] Wald R and Yurtsever U 1991 General proof of the averaged null energy condition for a massless scalar field in two-dimensional curved spacetime *Phys. Rev. D* **44** 403–16
- [12] Fewster C J and Galloway G J 2011 Singularity theorems from weakened energy conditions *Class. Quantum Grav.* **28** 125009
- [13] Brown P J, Fewster C J and Kontou E-A 2018 A singularity theorem for Einstein–Klein–Gordon theory *Gen. Relativ. Gravit.* **50** 121
- [14] Ehlers J 1993 Contributions to the relativistic mechanics of continuous media *Gen. Relativ. Gravit.* **25** 1225–66
Ehlers J 1961 Contributions to the relativistic mechanics of continuous media *Abh. Akad. Wiss. Lit. Mainz. Nat.* **K1.11** 793
- [15] Fewster C J and Kontou E-A 2020 A new derivation of singularity theorems with weakened energy hypotheses *Class. Quantum Grav.* **37** 065010
- [16] O’Neill B 1983 *Semi-Riemannian Geometry with Applications to Relativity (Pure and Applied Mathematics vol 103)* (New York: Academic)
- [17] Kriele M 1999 *Spacetime: Foundations of General Relativity and Differential Geometry (Lecture Notes in Physics. New Series m: Monographs vol 59)* (Berlin: Springer)

- [18] Galloway G J and Senovilla J M M 2010 Singularity theorems based on trapped submanifolds of arbitrary co-dimension *Class. Quantum Grav.* **27** 152002
- [19] Aazami A B and Javaloyes M A 2016 Penrose's singularity theorem in a Finsler spacetime *Class. Quantum Grav.* **33** 025003
- [20] Schinnerl B and Steinbauer R 2021 A note on the Gannon–Lee theorem *Lett. Math. Phys.* **111** 142
- [21] Ehrlich P E and Kim S-B 2009 Comparison theory in Lorentzian and Riemannian geometry *Nonlinear Anal. Theory Methods Appl.* **71** 211–7
- [22] Ford L H 1978 Quantum coherence effects and the second law of thermodynamics *Proc. R. Soc. A* **364** 227–36
- [23] Ford L H 1991 Constraints on negative-energy fluxes *Phys. Rev. D* **43** 3972–8
- [24] Fewster C J 2017 Quantum energy inequalities *Wormholes Warp Drives and Energy Conditions (Fundamental Theories of Physics vol 189)* ed F Lobo (Berlin: Springer) pp 215–54
- [25] Fewster C J and Kontou E-A 2019 Quantum strong energy inequalities *Phys. Rev. D* **99** 045001
- [26] Parker L and Fulling S A 1973 Quantized matter fields and the avoidance of singularities in general relativity *Phys. Rev. D* **7** 2357–74
- [27] Fischetti M V, Hartle J B and Hu B L 1979 Quantum effects in the early universe: I. Influence of trace anomalies on homogeneous, isotropic, classical geometries *Phys. Rev. D* **20** 1757–71
- [28] Anderson P 1983 Effects of quantum fields on singularities and particle horizons in the early universe *Phys. Rev. D* **28** 271–85
- [29] Anderson P 1983 Erratum: effects of quantum fields on singularities and particle horizons in the early universe *Phys. Rev. D* **28** 2695
- [30] Barcelo C, Liberati S, Sonogo S and Visser M 2008 Fate of gravitational collapse in semiclassical gravity *Phys. Rev. D* **77** 044032
- [31] Hayward S A 2006 Formation and evaporation of nonsingular black holes *Phys. Rev. Lett.* **96** 031103
- [32] Aghanim N *et al* 2018 Planck 2018 results: VI. Cosmological parameters (arXiv:1807.06209)
- [33] Misner C W, Thorne K and Wheeler J 1973 *Gravitation* (San Francisco, CA: Freeman)
- [34] Khavkine I and Moretti V 2015 Algebraic QFT in curved spacetime and quasifree Hadamard states: an introduction *Advances in Algebraic Quantum Field Theory (Mathematical Physics Studies)* (Berlin: Springer) pp 191–251
- [35] Moretti V 2021 On the global Hadamard parametrix in QFT and the signed squared geodesic distance defined in domains larger than convex normal neighbourhoods *Lett. Math. Phys.* **111** 130
- [36] Kay B S and Wald R M 1991 Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon *Phys. Rep.* **207** 49–136
- [37] Hollands S and Wald R M 2001 Local Wick polynomials and time ordered products of quantum fields in curved space-time *Commun. Math. Phys.* **223** 289–326
- [38] Fewster C J and Verch R 2012 Dynamical locality and covariance: what makes a physical theory the same in all spacetimes? *Ann. Henri Poincaré* **13** 1613–74
- [39] Fewster C J 2018 The art of the state *Int. J. Mod. Phys. D* **27** 1843007
- [40] Hollands S and Wald R M 2005 Conservation of the stress tensor in perturbative interacting quantum field theory in curved spacetimes *Rev. Math. Phys.* **17** 227–311
- [41] Sanders K 2021 Static symmetric solutions of the semi-classical Einstein–Klein–Gordon system *Ann. Henri Poincaré* (online first) (arXiv:2007.14311 [math-ph])
- [42] Pinamonti N and Siemssen D 2015 Global existence of solutions of the semiclassical Einstein equation for cosmological spacetimes *Commun. Math. Phys.* **334** 171–91
- [43] Gottschalk H and Siemssen D 2021 The cosmological semiclassical Einstein equation as an infinite-dimensional dynamical system *Ann. Henri Poincaré* **22** 3915–64
- [44] Fewster C J and Pfenning M J 2006 Quantum energy inequalities and local covariance: I. Globally hyperbolic spacetimes *J. Math. Phys.* **47** 082303
- [45] Eveson S P and Fewster C J 2007 Mass dependence of quantum energy inequality bounds *J. Math. Phys.* **48** 093506
- [46] Kontou E-A and Olum K D 2015 Quantum inequality in spacetimes with small curvature *Phys. Rev. D* **91** 104005
- [47] Connor S B and Fewster C J 2022 Integrals of incomplete beta functions, with applications to order statistics, random walks and string enumeration *Braz. J. Probab. Stat.* **36** 185–98
- [48] Fewster C J and Verch R 2002 A quantum weak energy inequality for Dirac fields in curved space-time *Commun. Math. Phys.* **225** 331–59

- [49] Dawson S P and Fewster C J 2006 An explicit quantum weak energy inequality for Dirac fields in curved spacetimes *Class. Quantum Grav.* **23** 6659–81
- [50] Fewster C J and Smith C J 2008 Absolute quantum energy inequalities in curved spacetime *Ann. Henri Poincaré* **9** 425–55
- [51] Fewster C J and Roman T A 2003 Null energy conditions in quantum field theory *Phys. Rev. D* **67** 044003
Fewster C J and Roman T A 2009 Null energy conditions in quantum field theory *Phys. Rev. D* **80** 069903 (erratum)
- [52] Freivogel B, Kontou E-A and Krommydas D 2020 The return of the singularities: applications of the smeared null energy condition (arXiv:2012.11569 [gr-qc])
- [53] Freivogel B and Krommydas D 2018 The smeared null energy condition *J. High Energy Phys.* **JHEP12(2018)067**
- [54] Meda P, Pinamonti N and Siemssen D 2021 Existence and uniqueness of solutions of the semiclassical Einstein equation in cosmological models *Ann. Henri Poincaré* **22** 3965–4015
- [55] Olver F W J, Olde Daalhuis A B, Lozier D W, Schneider B I, Boisvert R F, Clark C W, Miller B R and Saunders B V 2019 NIST digital library of mathematical functions (Release 1.0.22) <http://dlmf.nist.gov/> (accessed of 15 March 2019)
- [56] Muckenhoupt B 1972 Hardy’s inequality with weights *Stud. Math.* **44** 31–8
- [57] Kufner A, Kuliev K and Persson L-E 2012 Some higher order Hardy inequalities *J. Inequalities Appl.* **2012** 69
- [58] Kampé de Fériet J 1921 Les fonctions hypergéométriques d’ordre supérieur à deux variables *C. R. Acad. Sci. Paris* **173** 401–4
- [59] Kampé de Fériet J 1921 Quelques propriétés des fonctions hypergéométriques d’ordre supérieur à deux variables *C. R. Acad. Sci. Paris* **173** 489–91