Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Lectures on quantum field theory in curved spacetime

by

Christopher J. Fewster

Lecture note no.: 39 2008

Lectures on quantum field theory in curved spacetime

Christopher J. Fewster*
Department of Mathematics, University of York,
Heslington, York, YO10 5DD, UK

October 14, 2008

Abstract

These notes provide an introduction to quantum field theory in curved spacetimes, starting from the beginning but leading to some areas of current research. Topics covered include: globally hyperbolic spacetimes, canonical quantization, Euclidean Green functions, the Unruh effect, gravitational particle production, algebraic quantization, the Hadamard and microlocal spectrum conditions, and quantum energy inequalities.

*Electronic address: `cjf3@york.ac.uk`

Contents

1 Introduction, scope and literature

Quantum field theory in curved spacetime (QFT in CST) is the study of quantum fields moving in fixed curved spacetime backgrounds. One could think of this as a description of freely falling 'test quantum fields', just as geodesics describe the motion of freely falling 'test particles': the dynamics of the test system responds to the curvature of spacetime field, but the system does not modify the spacetime itself. So we are ignoring half of Wheeler's famous slogan that 'matter tells spacetime how to curve; spacetime tells matter how to move'.

Among the reasons for studying the theory are:

- QFT experiments at CERN, DESY etc are performed in a lightly curved spacetime background. If QFT could not be formulated in such an environment we could not claim to have understood terrestrial accelerator physics.
- The early universe and extreme astrophysical environments are far from the flat Minkowksi spacetime of conventional QFT, but should not need a full theory of quantum gravity for their description.
- The theory makes spectacular predictions such as radiation by black holes, and provides surprises (e.g., the Unruh effect)
- QFT in CST presents the challenge of understanding what structures and concepts are really important in QFT and which are merely useful simplifying assumptions.
- The theory can be sucessfully brought under mathematical control (at least for free fields and perturbation theory based thereon) using mathematical tools that are of interest in their own right.
- In the background, one also hopes that by clearly understanding QFT in CST one gains insights and intuitions into the interactions between gravitation and quantum theory with a view to quantum gravity proper, or QFT on structures other than manifolds [graphs, fractals, noncommutative spacetimes...].

This course will focus much more on the structures of QFT in CST than on applications. It will also concern a single quantum field model, the free scalar field, although the techniques described in the last lecture (section 6) form the foundations for the recent progress made by Fredenhagen & Brunetti and Hollands & Wald in formulating perturbation theory in curved spacetime.

The principal monographs on QFT in CST are, in chronological order:

- ND Birrell and PCW Davies, *Quantum fields in curved space* [6]
- SA Fulling, *Aspects of quantum field theory in curved space-time* [33]
- RM Wald, *Quantum field theory in curved spacetime and black hole thermodynamics* [71]

and the intention is that these lectures will make contact with each of these presentations at various stages. To a large extent they are complementary in terms of mathematical style and content; however, none of them covers the material presented in the later sections of these notes. The recently published

- C Bär, N Ginoux and F Pfäffle, *Wave equations on Lorentzian manifolds and quantization* [2]

provides a valuable summary of the classical PDE theory (local and global) relevant to the subject. The course begins with a rather rapid summary of the differential geometry relevant to the theory, for which standard advanced GR texts such as

- SW Hawking and GFR Ellis, *The large scale structure of space-time* [38]
- RM Wald, *General relativity* [70]
- B O'Neill, *Semi-Riemannian geometry* [57]

(the last of which is more pure mathematical in tone) give more leisurely and elegant treatments.

Microlocal analysis makes an appearance in section 6, and

- L Hörmander, *The analysis of linear partial differential operators I* [44]

gives an excellent and thorough account. A number of the original papers that apply the theory to QFT in CST have accessible summaries of the relevant portions of the theory. Regarding references: in a contribution of this scope one cannot hope to be comprehensive and the bibliography is biased towards papers written after the monographs given above, which themselves contain thorough bibliographies of earlier material.

Finally, although there are theorems to back up (hopefully) everything I say, I do not dwell on proofs. The aim is also to try to indicate why the theory is the way it is, rather than simply to present it from the start as a pristine mathematical structure.

Notes and acknowledgments: The material contained in sections 2-6.3 was given as five 90-minute lectures over the space of three days at the Spring School on Quantum Structures held at the University of Leipzig in February 2008. Section 6.4 was covered as part of a concluding seminar session, while sections 1 and 2 were distributed in advance as preliminary background reading. (I was asked not to assume that all participants would have a background in general relativity.) I would like to thank the organisers and participants of the Spring School for providing the opportunity and a stimulating environment in which to give these lectures. Thanks are also due to Falk Lindner for creating the figures from my blackboard sketches and typing parts of section 6, Bernard Kay for useful discussions relating to section 6 and to Ko Sanders for useful comments on sections of the notes.

2 Manifolds, covariant derivatives and all that

We summarise the main geometric concepts to be used in the course. This section can be skimmed or skipped by those conversant with GR or differential geometry.

2.1 Vectors, covectors...

Recall some basic multivariable calculus: if $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth function, then $D\varphi|_x$ denotes its derivative at x, which is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ obeying

$$D\varphi|_x(v) = \lim_{t\to 0} \frac{\varphi(x+tv) - \varphi(x)}{t}$$

for all $v \in \mathbb{R}^n$. The chain rule reads

$$D(\varphi \circ \psi)|_x = D\varphi|_{\psi(x)} D\psi|_x.$$

I assume some basic familiarity with smooth manifolds so much of this section is intended as revision rather than a systematic development. Thus I expect you to know that a **coordinate patch** or **chart** in M is an open subset $U \subset M$ with a homeomorphism κ mapping U to an open subset of $\mathbb{R}^n$, and that if (U, κ) and (U', κ') are coordinate patches with $U \cap U' \neq \emptyset$ then there is a smooth transition function $\varphi : \kappa(U \cap U') \to \kappa'(U \cap U')$, with $\varphi = \kappa' \circ \kappa^{-1}$. Although not the most elegant thing to do, geometric objects on M may be represented using coordinate patches. Where patches overlap the same object will have a number of different 'chart expressions' and the way that these depend on the transition functions is known as the transformation law.

Functions $f : M \to \mathbb{R}$ have chart expressions $f_\kappa : \kappa(U) \to \mathbb{R}$ given by $f_\kappa = f \circ \kappa^{-1}$ and transform as **scalars**

$$f_{\kappa'}(\varphi(x)) = f_\kappa(x).$$

Vector fields X on M have chart expressions $X_\kappa : \kappa(U) \to \mathbb{R}^n$ (thought of as column vectors) and transform **contravariantly**

$$X_{\kappa'}(\varphi(x)) = D\varphi|_x X_\kappa(x).$$

Covector fields ξ on M have chart expressions $\xi_\kappa : \kappa(U) \to (\mathbb{R}^n)^*$ (thought of as row vectors) and transform **covariantly**

$$\xi_{\kappa'}(\varphi(x)) D\varphi|_x = \xi_\kappa(x).$$

A **density** ρ of weight k is an object with chart expressions $\rho_\kappa : \kappa(U) \to \mathbb{R}$, transforming as

$$\rho_{\kappa'}(\varphi(x)) |\det D\varphi|_x|^k = \rho_\kappa(x).$$

Examples

1. If $c : \mathbb{R} \to M$ is a smooth curve, it has a tangent vector $\dot{c}(t)$ at $c(t)$ given [in any chart near $c(t)$] by

$$\dot{c}(t)_\kappa = D(\kappa \circ c)|_t = \frac{d}{dt}\kappa(c(t))$$

and the chain rule shows that this transforms contravariantly between charts:

$$\dot{c}(t)_{\kappa'} = D(\kappa' \circ c)|_t = D(\varphi \circ \kappa \circ c)|_t = D\varphi|_{\kappa(c(t))} D(\kappa \circ c)|_t = D\varphi|_{\kappa(c(t))}\dot{c}(t)_\kappa.$$

Similarly, if $f : M \to \mathbb{R}$ is a smooth function, there is a covector field ∇f given by

$$(\nabla f)_\kappa = Df_\kappa$$

in each chart and the chain rule gives

$$(\nabla f)_\kappa = D(f_{\kappa'} \circ \varphi) = Df_{\kappa'} D\varphi = (\nabla f)_{\kappa'} D\varphi$$

i.e., covariant transformation.

2. Densities of weight 1 are often called just 'densities', and the transformation property is exactly right to ensure that

$$\int_{\kappa(U)} \rho_\kappa(x) d^n x = \int_{\kappa'(U)} \rho_{\kappa'}(y) d^n y$$

for a density supported in $U \cap U'$. Thus it makes sense to define the integral of ρ by

$$\int_M \rho = \int_{\kappa(U)} \rho_\kappa(x) d^n x$$

if ρ is supported in U, using a partition of unity to define the integral of a general density. Note that a density 'contains its own integration measure'.

3. If X is a vector field and ξ a covector field, there is a (scalar) function $\xi(X)$ such that

$$\xi(X)_\kappa = \xi_\kappa \cdot X_\kappa,$$

where the $\cdot$ is just ordinary matrix multiplication of row and column vectors.

As a digression, it should be noted that there are much tidier formulations of the concepts just described (see, e.g., [70]). For example, a vector at p may be characterised as a linear map v from $C^\infty(M; \mathbb{R})$ to $\mathbb{R}$ with the Leibniz property

$$v(fg) = v(f)g(p) + f(p)v(g) \qquad f, g \in C^\infty(M; \mathbb{R}).$$

Or, yet again, vectors at p may be characterised as the equivalence classes of curves through p under the equivalence

$$c \sim c' \iff D(f \circ c)(0) = D(f \circ c')(0)$$

assuming $c(0) = p = c'(0)$.

2.2 Tensors and index notation

The set of all vectors at p forms a vector space T_pM, the tangent space to M at p. Its dual space T_p^*M may be identified with the vector space of covectors at p, as the last example shows.

A **tensor** of type (k, l) at p is a multilinear map from a Cartesian product of k copies of T_p^*M and l copies of T_pM to $\mathbb{R}$. Equivalently, it belongs to a tensor product of k copies of T_pM and l copies of T_p^*M. In particular, any vector is a type $(1, 0)$-tensor and a covector is of type $(0, 1)$.

We use an **abstract index notation** to keep track of calculations with tensors. This is best illustrated by an example: an expression $S^a{}_b{}^c$ denotes a $(2, 1)$-tensor, which (at some $p \in M$) we regard as a multilinear map $S : T_p^*M \times T_pM \times T_p^*M \to \mathbb{R}$, or (equivalently) as an element of $T_pM \otimes T_p^*M \otimes T_pM$. If $\xi, \eta \in T_p^*M$ and $v \in T_pM$, we use the notation $S^a{}_b{}^c \xi_a v^b \eta_c$ for $S(\xi, v, \eta)$; similarly, the notation $S^a{}_b{}^c \xi_a v^b$ denotes the linear map

$$T_p^*M \ni \eta \mapsto S(\xi, v, \eta) \in \mathbb{R},$$

which can be identified canonically with an element of T_pM, i.e., a vector. The indices are not 'doing' anything beyond keeping track of the slots; a paired index denotes a substituted slot, while the unpaired indices in a given term give its overall tensor type. However the abstract index notation also mirrors what is happening in charts, where tensors become arrays of components and a pairing implements a sum. We use Greek indices for these purposes, and with this convention it is now safe to suppress chart subscripts if we use e.g., primed indices where we might have used a κ'.

Example Let S be a smooth $(0, 2)$-tensor field. Its chart expression is an array $S_{\alpha\beta}$, and the transformation property is

$$S_{\alpha'\beta'} (D\varphi|_x)^{\alpha'}{}_\alpha (D\varphi|_x)^{\beta'}{}_\beta = S_{\alpha\beta}.$$

As a consequence, we obtain a (possibly nonsmooth) density ρ by setting

$$\rho_\kappa = \sqrt{|\det S_{\alpha\beta}|}$$

for each chart κ. We say that S is **nondegenerate** (at p) if $S_{ab} u^a v^b = 0$ for all u (at p) implies that $v = 0$. If this condition is satisfied the determinant will be nonvanishing and a smooth density can be obtained by the above method.

2.3 Metric and covariant differentiation

A **metric** is an everywhere smooth, nondegenerate, type $(0, 2)$-tensor field that is also symmetric, i.e., $S_{ab} = S_{ba}$ (equivalently $S(u, v) = S(v, u)$ for all u and v). The above discussion shows that this immediately induces a density ρ_g, which we call the associated volume element and is often written $d\mathrm{vol}_g$. The metric is often stated in a component form as

$$ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta.$$

Given a metric, we have a definition of integration on the manifold by

$$\int_M f(p) d\mathrm{vol}_g(p) = \int_M f\rho_g.$$

The signature of the metric at p is the difference between the numbers of positive and negative eigenvalues in its chart expression at $\kappa(p)$; this is invariant under coordinate transformations. We work with **Lorentzian metrics**, which [in our convention] have signature $2-n$ on manifolds of dimension n. Vectors with $g(u,u) > 0$ are classified as **timelike**; those with $g(u,u) < 0$ as **spacelike**; those with $g(u,u) = 0$ as **null**. A curve whose tangent vector is everywhere timelike (resp., null; spacelike) is said to be timelike (resp., null; spacelike); if a curve is nowhere spacelike it is said to be **causal**.

Owing to its nondegeneracy, the metric induces isomorphisms between the spaces of vectors and covectors at a given point. If u is a vector, $u^\flat$ is the unique covector such that

$$u^\flat(v) = g(u,v)$$

for all vectors v. In index notation, $(u^\flat)_a = g_{ab}u^b$ and we usually dispense with the $\flat$, simply writing

$$u_a = g_{ab}u^b.$$

The inverse map to $\flat$ is of course denoted $\sharp$, so to each covector ξ there is a uniquely defined vector $\xi^\sharp$ with $\xi^{\sharp\flat} = \xi$. Again, we typically write $(\xi^\sharp)^a$ simply as ξ^a. This determines a symmetric type $(2,0)$-tensor g^{ab} with the defining property

$$\xi^a = g^{ab}\xi_a$$

for all covectors ξ, which obeys

$$g^{ab}g_{bc} = \delta^a{}_c$$

(the chart expression of the right-hand side is the identity matrix in any chart). For obvious reasons, these various procedures are known as raising and lowering indices.

A metric also introduces a notion of **covariant derivative**. This extends the operator ∇ from functions to tensors of arbitrary type. There is a unique way of doing this, subject to the conditions:

- $\nabla_a g_{bc} = 0$
- $\nabla_a\nabla_b f = \nabla_b\nabla_a f = 0$ for all *functions* f
- Leibniz' rule holds.

In a chart we have

$$\nabla_\beta u^\alpha = u^\alpha{}_{,\beta} + \Gamma^\alpha_{\beta\gamma}u^\gamma,$$

where the $_{,\beta}$ indicates partial derivative with respect to the β'th coordinate and the Christoffel symbols are

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2}g^{\alpha\delta}\left(g_{\beta\delta,\gamma} + g_{\gamma\delta,\beta} - g_{\beta\gamma,\delta}\right).$$

From this and the Leibniz rule, we can differentiate any tensor: for instance, $\nabla_a \xi_b$ is determined by the condition that, for any vector u,

$$u^b \nabla_a \xi_b = \nabla_a(\xi_b u^b) - \xi_b \nabla_a u^b$$

and the right-hand side consists of derivatives we already know how to do.

Commutativity of second derivatives only holds for scalars in general; its failure for vectors indicates the presence of **curvature**. To be precise, the **Riemann tensor** has the defining property that

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) v^d = R_{abc}{}^d v^c$$

(and vanishes in flat spacetime). We define the associated tensors

$$\begin{aligned} \text{Ricci tensor} \qquad & R_{ab} = R^d{}_{adb} \\ \text{Ricci scalar} \qquad & R = g^{ab} R_{ab} \\ \text{Einstein tensor} \qquad & G_{ab} = R_{ab} - \frac{1}{2} R g_{ab}. \end{aligned}$$

2.4 Geodesics

Let $c : \mathbb{R} \to M$ be a smooth curve, with tangent vector $\dot{c}(\lambda)$ at $c(\lambda)$. A vector field u defined near c is parallel-transported along c if

$$\dot{c}(\lambda)^a \nabla_a u^b|_{c(\lambda)} = 0$$

for all $\lambda \in \mathbb{R}$. If $u^a(c(\lambda)) = \dot{c}(\lambda)^a$, i.e., u extends $\dot{c}$, and u is parallel-transported, then we say that c is an **affinely parameterised geodesic**. In terms of coordinates, the requirement is

$$\ddot{c}^\alpha + \Gamma^\alpha_{\beta\gamma} \dot{c}^\beta \dot{c}^\gamma = 0.$$

This is also easily seen to be the Lagrange equation corresponding to Lagrangian

$$\mathscr{L}(x, u) = \frac{1}{2} g_{\alpha\beta}(x) u^\alpha u^\beta,$$

which is also a constant of the motion [no explicit λ-dependence]. Hence a geodesic which has timelike tangent vector at one point must have an everywhere timelike tangent; likewise for null and spacelike geodesics. In general relativity, a freely falling massive test particle follows a geodesic with timelike tangent vector [i.e., a timelike geodesic], while a freely falling massless particle follows a null geodesic.

In Minkowski spacetime there is a unique geodesic between any two points, but this is not always true on general manifolds; there may be multiple geodesics between points (e.g., antipodal points on a sphere) or points with no geodesic between them (e.g., Minkowski space with a point removed). However,[1] every point in a semi-Riemannian manifold has a **convex normal neighbourhood**, in which any two points may be joined by *a unique geodesic lying in that neighbourhood* [there may be others which leave it].

[1] See, e.g., Prop. 5.7 in [57].

2.5 General relativity

This is not a course on GR, and one could simply view QFT in CST simply as a physical theory formulated on manifolds, though this would rather lose sight of its origins. For those who have not seen them before, the field equations for general relativity are Einstein's equations:

$$G_{ab} + \Lambda g_{ab} = -8\pi G T_{ab},$$

where T_{ab} is the stress-energy tensor of the energy-matter content of spacetime, Λ is the cosmological constant and G is Newton's constant. As the Einsten tensor is symmetric and **conserved**, i.e., $\nabla^a G_{ab} = 0$, we may deduce that T_{ab} must also be symmetric and conserved if it is to stand on the right-hand side of the Einstein equations.

In particular models one can obtain T_{ab} from a matter action S_{m},

$$T_{ab} = \frac{2}{\rho_g} \frac{\delta S_{\mathrm{m}}}{\delta g^{ab}}$$

and indeed the Einstein equations may then be obtained by demanding that $S_{\mathrm{g}} + \kappa S_{\mathrm{m}}$ be stationary with respect to variations of g^{ab}, where

$$S_{\mathrm{g}} = \frac{1}{16\pi G} \int \rho_g (R - 2\Lambda).$$

(More precisely, we should work locally with convergent integrals, as we do for the Klein–Gordon field in section 3.1; we have also ignored questions relating to boundaries).

2.6 Summary of main conventions

- The metric has signature $+ - - \cdots$
- $(\nabla_a \nabla_b - \nabla_b \nabla_a) v^d = R_{abc}{}^d v^c$, and $R_{ab} = R^d{}_{adb}$.
- Latin indices denote abstract indices; Greek indices denote coordinate components.
- Henceforth $\hbar = c = G = 1$.
- Fourier transforms will be nonstandardly defined by

  $$\widehat{f}(k) = \int d^n x \, e^{ik\cdot x} f(x);$$

  the hat will sometimes be displaced e.g., $f^\wedge(k)$, for typographical reasons.

3 The classical Klein–Gordon field

The main focus of this course will be the quantum field theory of the real scalar field. This reflects both the relative simplicity of the model and its predominance in the literature on QFT in CST. We begin by introducing the classical Klein–Gordon equation, and describing a class of spacetimes on which it is well-posed. Passing to a Hamiltonian formulation, we identify suitable classical observables that will form the basis for the quantization of the theory according to Dirac's prescription.

3.1 The Klein–Gordon action

The Klein–Gordon theory on a spacetime (M, g) is the field theory defined by the Lagrangian density

$$\mathscr{L}_g[\phi] = \frac{1}{2}\rho_g \left(g^{ab}(\nabla_a\phi)(\nabla_b\phi) - (m^2 + \xi R)\phi^2\right),$$

where ϕ is a scalar field. The main differences from the flat action are that we have g^{ab} instead of η^{ab}, and have made allowance for an extra term $\xi R\phi^2$, where R is the Ricci scalar and ξ is a dimensionless coupling constant. This extra term is added partly because it simply has the correct dimensions, and partly because for the special value $\xi = (n-2)/(4n-4)$ the action exhibits conformal invariance in the massless case $m = 0$, meaning that the Lagrangian density is unchanged under the simultaneous replacements

$$g_{ab} \to \overline{g}_{ab} = \Omega^2 g_{ab} \qquad \phi \to \overline{\phi} = \Omega^{1-n/2}\phi$$

for any smooth positive function Ω, i.e., $\mathscr{L}_{\overline{g}}[\overline{\phi}] = \mathscr{L}_g[\phi]$.

If one studies perturbative QFT based on the free Klein–Gordon theory then the term $\xi R\phi^2$ will enter as a counter-term so it makes sense to be able to deal with it from the start. The case $\xi = 0$ is known as **minimal coupling** and $\xi \neq 0$ as **non-minimal coupling**, with the special values $\xi = (n-2)/(4n-4)$ known unsurprisingly as **conformal coupling**.

The action formed from $\mathscr{L}$ is

$$S[U; \phi] = \int_U \mathscr{L}_g[\phi],$$

where U is a compact submanifold of M, and the equation of motion is obtained by demanding that for every such U, $S[U; \phi]$ is stationary with respect to smooth variations of ϕ that vanish on ∂U. This results in the Klein–Gordon equation

$$P\phi := (\Box_g + m^2 + \xi R)\phi = 0,$$

where

$$\Box_g = g^{ab}\nabla_a\nabla_b.$$

In coordinates we may write

$$\Box_g = g^{\alpha\beta}(\partial_\alpha\partial_\beta - \Gamma^\gamma_{\alpha\beta}\partial_\gamma),$$

or the rather nicer form

$$\Box_g = \frac{1}{\sqrt{-g}}\partial_\alpha g^{\alpha\beta}\sqrt{-g}\partial_\beta,$$

which makes clear that

$$\int_M d\mathrm{vol}_g \phi \Box_g f = -\int_M d\mathrm{vol}_g g^{ab}(\nabla_a\phi)(\nabla_b f) = \int_M d\mathrm{vol}_g (\Box_g\phi) f$$

for any $\phi \in C^\infty(M)$ and $f \in C_0^\infty(M)$.

Finally, the stress-energy tensor is obtained by varying the action with respect to the metric, which gives

$$\begin{aligned} T_{ab} &= (\nabla_a\varphi)(\nabla_b\varphi) - \frac{1}{2}g_{ab}g^{cd}(\nabla_c\varphi)(\nabla_d\varphi) + \frac{1}{2}m^2 g_{ab}\varphi^2 \\ &\quad + \xi\left(g_{ab}\Box_g - \nabla_a\nabla_b - G_{ab}\right)\phi^2, \end{aligned}$$

where G_{ab} is the Einstein tensor. Note that the effect of the coupling constant can be seen in the stress-energy tensor even where the metric is Ricci flat, even though the $\xi R\phi^2$ term in the Klein–Gordon equation vanishes in such situations.

3.2 Global hyperbolicity

The existence of solutions to the Klein–Gordon equation is quite sensitive to the global geometry and topology of spacetime. We work with the class of **globally hyperbolic spacetimes** where everything works nicely. We assume that the spacetime is connected and **time-oriented**, i.e., it admits a vector field u with $g_{ab}u^a u^b > 0$ everywhere, permitting us to classify any timelike or null vector v^a as future-pointing [if $u_a v^a > 0$] or past-pointing [if $u_a v^a < 0$].

Definition 3.1 *A spacetime* (M, g) *is* globally hyperbolic *if it admits a* Cauchy surface *i.e., a subset intersected exactly once by every inextendible timelike curve in* (M, g)*.*

Lemma 14.29 in [57] shows that any Cauchy surface is in fact a closed achronal[2] topological hypersurface[3] met exactly once by every inextendible causal curve in (M, g).

Examples: Minkowski space (M_0, η) is globally hyperbolic, with respect to the Cauchy surface $t = 0$; so is the **Rindler wedge**

$$R = \{(t, x, y, z) \in M_0 : z > |t|\}$$

with the induced metric from (M_0, η) again with respect to the portion of $t = 0$ lying in R. Further examples are sketched in Figure 1.

Global hyperbolicity rules out pathologies such as (i) closed or almost-closed causal curves; (ii) causal curves that can 'fall off the edge' of spacetime in finite coordinate time.

[2] No timelike curve meets it more than once.

[3] Locally homeomorphic to a hyperplane in $\mathbb{R}^n$

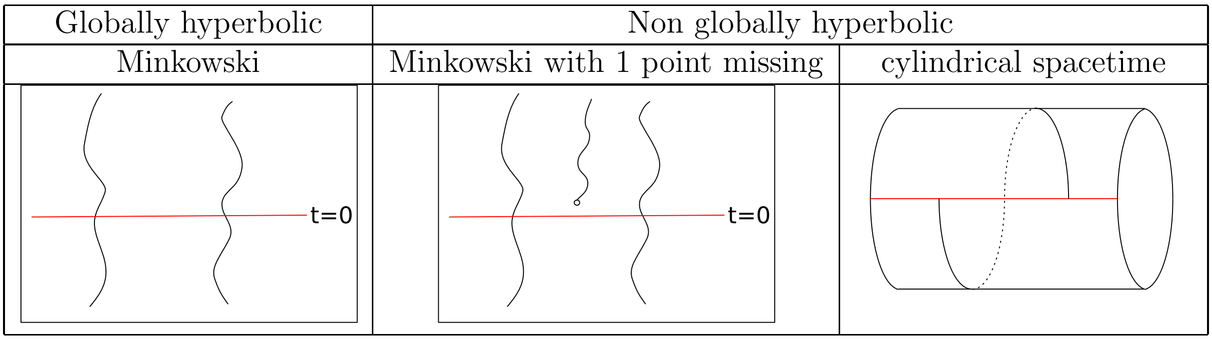


Figure 1: Examples of (non) globally hyperbolic spacetimes. The wavy lines in the first two panels are timelike curves. In Minkowski space, all timelike curves (can be extended to) cross the Cauchy surface $t = 0$. However if a point is removed there will be some timelike curves that do not have this property and the spacetime is no longer globally hyperbolic. In the final example we periodically identify Minkowski space under $t \mapsto t + T$ to obtain the 'spacelike cylinder' spacetime, which contains timelike curves that cut the image of the $t = 0$ surface more than once.

The first of these is known as the strong causality condition, while the second of these can be more precisely stated as follows: if p and q are connected by a future-directed causal curve, then

$$J^+(p) \cap J^-(q)$$

is compact, where $J^\pm(p)$ denotes all points that can be reached by future(+)/past(−)-directed causal curves from p [including p itself]. Indeed, global hyperbolicity is equivalent to the requirement of causality (no closed causal curves) plus compactness of all subsets $J^+(p) \cap J^-(q)$, and implies strong causality (Theorem 3.2 in [5]).

A long-standing conjecture, finally resolved quite recently by Bernal and Sánchez, is that any globally hyperbolic spacetime can be smoothly foliated into Cauchy surfaces. In fact, their result shows that this can be done in a particularly nice way:

Theorem 3.2 (see Theorem 1.2 in [4] and the proof of Prop. 2.4 in [3]) *Let S be a smooth spacelike Cauchy surface in a globally hyperbolic spacetime (M, g). Then there is an isometry of (M, g) to the smooth product manifold $\mathbb{R} \times S$ with metric $N^2 dt^2 - h_t$, so that*

- *N is smooth and positive,*
- *each h_t is a (smooth) Riemannian metric on S with $t \mapsto h_t$ smooth,*
- *each $\{t\} \times S$ is a smooth spacelike Cauchy surface.*

For our current purposes the main implication of global hyperbolicity is that the Klein–Gordon system is well-posed.

Theorem 3.3 *If (M, g) is globally hyperbolic there exist continuous maps $E^\pm : C_0^\infty(M) \to C^\infty(M)$ so that, for each $f \in C_0^\infty(M)$, $\phi = E^\pm f$ solves the inhomogeneous problem*

$$P\phi = f, \tag{1}$$

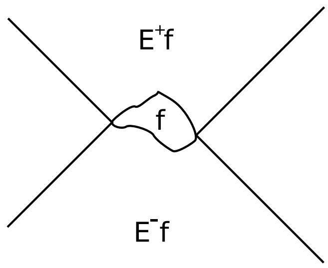


Figure 2: The support of the solutions $E^{\pm}f$.

has support $\mathrm{supp}\,\phi \subset J^{\pm}(\mathrm{supp}\,f)$*, and is the unique solution to* (1) *whose support has compact intersection with* $J^{\mp}(\mathrm{supp}\,f)$*.*

Due to the support properties (illustrated in Fig. 2), E^- (resp., E^+) is called the **advanced** (resp., **retarded**) **fundamental solution** (or Green function). In the special case where $f = Pf'$ for some $f' \in C_0^\infty(M)$, we note that f' has compact intersection with $J^{\mp}(\mathrm{supp}\,f)$ and so therefore $f' = E^{\pm}f$ by uniqueness. Hence we have

$$E^{\pm}Pf' = f'$$

in addition to the original property $PE^{\pm}f = f$.

The **advanced-minus-retarded fundamental solution** E is defined[4] by $E = E^- - E^+$. Clearly $\phi = Ef$ is a smooth solution to the homogeneous equation $P\phi = 0$, but we also have (cf. Theorem 3.4.7 in [2])

Theorem 3.4 *(i) Any smooth solution* ϕ *to* $P\phi = 0$ *which has compact support on some [and hence any] Cauchy surface may be written in the form*

$$\phi = Ef$$

for $f \in C_0^\infty(M)$*; given any neighbourhood* N *of a Cauchy surface one may choose such an* f *in* $C_0^\infty(N)$*. (ii) We have*

$$\ker E = PC_0^\infty(M).$$

In consequence, the space of smooth complex-valued solutions with compact support on Cauchy surfaces is

$$\mathsf{Sol} = EC_0^\infty(M) \cong C_0^\infty(M)/PC_0^\infty(M),$$

and the space of smooth real-valued solutions with compact support on Cauchy surfaces is

$$\mathsf{Sol}_{\mathbb{R}} = EC_0^\infty(M;\mathbb{R}) \cong C_0^\infty(M;\mathbb{R})/PC_0^\infty(M;\mathbb{R}).$$

[4]Warning: some authors use retarded-minus-advanced (e.g., [15]), or label retarded and advanced the other way round (e.g., [2])! Furthermore, in the $-+++$ signature, the fundamental solutions to $(\Box_g - m^2)\phi = f$ are *minus* the fundamental solutions we use; e.g., in [71], Wald's A is our $-E^-$.

We also freely regard E as a bidistribution, $E \in \mathscr{D}'(M \times M)$ so that

$$E(f_1, f_2) = \int_M d\mathrm{vol}_g(p) f_1(p)(Ef_2)(p)$$

for all $f_i \in C_0^\infty(M)$. Regarded like this, E is a weak bisolution:

$$E(Pf_1, f_2) = 0 = E(f_1, Pf_2)$$

for all $f_i \in C_0^\infty(M)$, with the antisymmetry property $E(f_1, f_2) = -E(f_2, f_1)$. It also vanishes except at causally related points, owing to the support properties of $E^\pm$.

Remark: Suppose (M, g) is globally hyperbolic and that $N \subset M$ is open and **causally convex**, i.e., any causal curve [in M] whose endpoints are contained in N lies completely in N. (The interior of any set $J^+(p) \cap J^-(q)$ clearly has this property.) Endowing N with the metric $g|_N$ and the time-orientation induced from (M, g), $(N, g|_N)$ obeys strong causality and compactness of its $J^+(p) \cap J^-(q)$'s *a fortiori*. So (N, g_N) becomes a globally hyperbolic spacetime in its own right. By uniqueness of solution it is then clear that $E^\pm_{(M,g)} f$ and $E^\pm_{(N,g|_N)} f$ must agree on N for all $f \in C_0^\infty(N)$; hence

$$E_{(M,g)}(f_1, f_2) = E_{(N,g|_N)}(f_1, f_2)$$

for $f_i \in C_0^\infty(N)$.

3.3 Canonical structure

The solution space for the real Klein–Gordon field, $\mathsf{Sol}_\mathbb{R}$, may be identified with the phase space of the theory. Classical observables are functions on this phase space: for example, every $f \in C_0^\infty(M; \mathbb{R})$ defines an observable F_f which acts on solutions $\phi \in \mathsf{Sol}_\mathbb{R}$ by

$$F_f(\phi) = \int_M d\mathrm{vol}_g(p) \phi(p) f(p).$$

We may observe that the F_f depend linearly on f, and that some of them vanish identically:

$$\begin{aligned} F_{Pf}(\phi) &= \int_M d\mathrm{vol}_g(p) \phi(p)((\Box_g + m^2 + \xi R) f)(p) \\ &= \int_M d\mathrm{vol}_g(p)((\Box_g + m^2 + \xi R)\phi)(p) f(p) \\ &= 0 \end{aligned}$$

for any $f \in C_0^\infty(M; \mathbb{R})$ and $\phi \in \mathsf{Sol}_\mathbb{R}$. Our main aim in this section will be to show that the Poisson bracket of two such observables is

$$\{F_{f_1}, F_{f_2}\} = E(f_1, f_2), \tag{2}$$

which will enable the quantization of the theory by means of Dirac's prescription.

To do this, we return to the Lagrangian density, and also suppose that spacetime has been foliated by Cauchy surfaces in the manner of Theorem 3.2. Thus we assume that (M, g) is isometric to $\mathbb{R} \times \Sigma$, with metric

$$ds^2 = N^2 dt^2 - (h_t)_{ij} dx^i dx^j,$$

where $N \in C_0^\infty(\mathbb{R} \times \Sigma; (0, \infty))$ is called the lapse function and h_t is a smooth Riemannian metric on Σ depending smoothly on t.

We can now describe the Klein–Gordon equation in terms of Lagrangian mechanics. Configuration space is the space of smooth real-valued functions on Σ, $C_0^\infty(\Sigma; \mathbb{R})$; each solution $\phi \in \mathsf{Sol}_{\mathbb{R}}$ may be regarded as a function

$$\begin{aligned} \mathbb{R} &\to C_0^\infty(\Sigma; \mathbb{R}) \\ t &\mapsto \varphi_t, \end{aligned}$$

where $\varphi_t(\underline{x}) = \phi(t, \underline{x})$. Letting U be the spacetime region bounded between surfaces $\{t_0\} \times \Sigma$ and $\{t_1\} \times \Sigma$, the action

$$S = \int_U \mathscr{L}_g[\phi]$$

may be written

$$S = \int_{t_0}^{t_1} dt L(\varphi_t, \dot{\varphi}_t, t),$$

where the dot denotes differentiation with respect to t and the Lagrangian is

$$L(\varphi, u, t) = \int_{\{t\}\times\Sigma} \frac{1}{2}\rho_h \left(N^{-1}u^2 - Nh^{ij}\nabla_i\varphi\nabla_j\varphi - N(m^2 + \xi R)\varphi^2 \right).$$

The momentum canonically conjugate to φ is a density[5]

$$\pi(\underline{x}, t) = \frac{\delta L}{\delta u(\underline{x})} = \frac{\rho_h(\underline{x})}{N(t, \underline{x})} u(\underline{x}),$$

and we pass to a Hamiltonian description by adopting (φ, π) as the basic variables on the phase space $\mathscr{P} = C_0^\infty(\Sigma; \mathbb{R}) \times \rho_h C_0^\infty(\Sigma; \mathbb{R})$, where the second factor is a shorthand for the space of smooth densities of compact support on Σ. The theory can then be described in terms of the dynamics arising from the Hamiltonian

$$H(\varphi, \pi, t) = \int_{\{t\}\times\Sigma} \frac{1}{2} N \left(\rho_h^{-1}\pi^2 + \rho_h h^{ij}\nabla_i\varphi\nabla_j\varphi + (m^2 + \xi R)\varphi^2 \right).$$

[5] $\delta L/\delta u(\underline{x})$ (understood to be evaluated at (φ, u)) is defined to be the unique density such that

$$\int_\Sigma \frac{\delta L}{\delta u(\underline{x})} f(\underline{x}) = \left. \frac{d}{d\lambda} L[\varphi, u + \lambda f] \right|_{\lambda=0}$$

for all smooth functions f.

Our main task is to calculate the Poisson brackets of the classical observables F_f. In general, Poisson brackets on $\mathscr{P}$ are defined so that

$$\{F, G\}(\varphi, \pi) = \int_\Sigma \left(\frac{\delta F}{\delta \varphi} \frac{\delta G}{\delta \pi} - \frac{\delta F}{\delta \pi} \frac{\delta G}{\delta \varphi} \right)$$

for differentiable $F, G : \mathscr{P} \to \mathbb{R}$. The right-hand side is well-defined, because functional derivatives with respect to the function φ are densities and those with respect to the density π are functions, so both terms of the integrand are densities.

Returning to our functions $F_f : \mathsf{Sol}_\mathbb{R} \to \mathbb{R}$, we use a useful standard identity[6]

$$F_f(\phi) = \int_M d\mathrm{vol}_g(p)\phi(p)f(p) = \int_\Sigma [\varphi_f(\underline{x})\pi(\underline{x}) - \pi_f(\underline{x})\varphi(\underline{x})], \tag{3}$$

where (φ, π) (resp., (φ_f, π_f)) is the phase space point corresponding to the solution ϕ (resp., Ef). It is clear from Eq. (3) that

$$\frac{\delta F_f}{\delta \varphi(\underline{x})} = -\pi_f(\underline{x}) \qquad \text{and} \qquad \frac{\delta F_f}{\delta \pi(\underline{x})} = \varphi_f(\underline{x}).$$

Setting $\phi = Ef'$ in the above, we then have

$$\begin{aligned} E(f, f') = F_f(Ef') &= \int_\Sigma [\varphi_f(\underline{x})\pi_{f'}(\underline{x}) - \pi_f(\underline{x})\varphi_{f'}(\underline{x})] \\ &= \int_\Sigma \left(\frac{\delta F_f}{\delta \varphi} \frac{\delta F_{f'}}{\delta \pi} - \frac{\delta F_f}{\delta \pi} \frac{\delta F_{f'}}{\delta \varphi} \right) \\ &= \{F_f, F_{f'}\}, \end{aligned}$$

as claimed in Eq. (2). Note that the construction of $\mathscr{P}$ involved a particular Cauchy surface, but the Poisson brackets of the F_f's are completely independent of the choice made.

Remark: To make contact with other treatments, we briefly mention that the real solution space $\mathsf{Sol}_\mathbb{R}$ can be equipped with a antisymmetric bilinear form $\sigma : \mathsf{Sol}_\mathbb{R} \times \mathsf{Sol}_\mathbb{R} \to \mathbb{R}$ defined by

$$\sigma(\phi, \phi') = \int_\Sigma [\varphi(\underline{x})\pi'(\underline{x}) - \pi(\underline{x})\varphi'(\underline{x})],$$

where (φ, π) and (φ', π') are the phase space points corresponding respectively to the solutions ϕ and ϕ'. The form σ is nondegenerate, in the sense that $\sigma(\phi, \phi')$ vanishes for all $\phi \in \mathsf{Sol}_\mathbb{R}$ if and only if $\phi' = 0$. In other words σ is a **symplectic form** on $\mathsf{Sol}_\mathbb{R}$. Our discussion above shows, among other things, that

$$\sigma(Ef, Ef') = E(f, f') \tag{4}$$

for all real-valued test-functions f, f', and this remains true for complex-valued test functions if σ is extended to a complex bilinear form on Sol.

[6] See, e.g., Lemma A.1 in [15], but note that the E used in that reference differs from ours by a sign.

3.4 Dirac quantization

Applying Dirac's quantization prescription to the classical observables F_f, we seek operators[7] $\widehat{F_f}$ ($f \in C_0^\infty(M)$) obeying

$$[\widehat{F_f}, \widehat{F_{f'}}] = i\{F_f, F_{f'}\}\mathbb{1} = iE(f, f')\mathbb{1} \tag{5}$$

and interpreted as smeared quantum fields. In particular, when the supports of f and f' are spacelike-separated, $\widehat{F_f}$ and $\widehat{F_{f'}}$ should commute, reflecting the Bose statistics of a spin-0 field. As a consequence of the commutation relations (5), we also have

$$[\widehat{\varphi}(p), \widehat{\pi}(f)] = i\int_\Sigma pf\,\mathbb{1},$$

where $\widehat{\varphi}(p)$ and $\widehat{\pi}(f)$ are quantizations of

$$\varphi(p) = \int_\Sigma \varphi(\underline{x})p(x) \qquad \pi(f) = \int_\Sigma \pi(\underline{x})f(\underline{x})$$

for smooth compactly supported density p and smooth compactly supported function f.

Recalling that $f \mapsto F_f$ is linear; that the F_f's are classical observables; and that $F_{Pf}(\phi) = 0$ for all $f \in C_0^\infty(M;\mathbb{R})$, $\phi \in \mathsf{Sol}_\mathbb{R}$, we would also expect that

- $f \mapsto \widehat{F_f}$ is real-linear;
- $\left(\widehat{F_f}\right)^* = \widehat{F_f}$ for all $f \in C_0^\infty(M;\mathbb{R})$;
- $\widehat{F_{Pf}} = 0$ for all $f \in C_0^\infty(M;\mathbb{R})$.

It is also convenient to permit smearings with complex-valued functions. Accordingly, we define

$$\widehat{\Phi}(f) = \widehat{F_{\mathrm{Re}\,f}} + i\widehat{F_{\mathrm{Im}\,f}}$$

for $f \in C_0^\infty(M)$ and seek to implement the following relations:

- $f \mapsto \widehat{\Phi}(f)$ is complex-linear;
- $\widehat{\Phi}(f)^* = \widehat{\Phi}(\overline{f})$ for all $f \in C_0^\infty(M)$;
- $\widehat{\Phi}(Pf) = 0$ for all $f \in C_0^\infty(M;\mathbb{R})$ for all $f \in C_0^\infty(M)$;
- $[\widehat{\Phi}(f), \widehat{\Phi}(f')] = iE(f, f')\mathbb{1}$ for all $f, f' \in C_0^\infty(M)$.

A key problem in QFT in CST is to find a Hilbert space on which $\widehat{\Phi}(f)$ satisfying the above relations may be defined. This can be done in two ways:

[7] We will write hats on top of operators only in this section. This should not be confused with the notation for a Fourier transform.

1. Direct constructions of the Hilbert space and smeared fields.
2. Splitting the problem into (a) first understanding the algebraic structure encoded in the above relations and then (b) the construction of Hilbert space representations of that algebra.

The second is representative of the algebraic approach to QFT in CST and will be a focus of the sections 5 and 6. First, however, we describe the direct approach in section 4.

4 Canonical quantization of the Klein–Gordon field

4.1 Quantization: the ultrastatic case

We say that (M, g_{ab}) is **ultrastatic** if there is a Riemannian manifold (Σ, h_{ij}) such that $M = \mathbb{R} \times \Sigma$ and $g = 1 \oplus -h$. We might write

$$ds^2 = dt^2 - h_{ij}(\underline{x})dx^i dx^j$$

where x^i denote coordinates on Σ and $\underline{x}$ a general point of Σ.

In these spacetimes it is particularly easy to follow canonical quantization methods and see what the quantum field theory might be. The aim is very much to follow our instinct and not to worry too much about details yet. The point of view we will adopt here is explained in much greater detail in Fulling's monograph [33] and is also influenced by the Fulling–Ruijsenaars paper [34].

The Klein–Gordon operator is

$$P = \frac{\partial^2}{\partial t^2} + K$$

where

$$K = -\triangle_h + (m^2 + \xi R)$$

and $\triangle_h = \rho_h^{-1}\partial_i \rho_h h^{ij}\partial_j$ is the Laplacian on (Σ, h). The operator K is symmetric on the domain $C_0^\infty(\Sigma)$ and we assume that it is extended to a self-adjoint operator, also denoted K. If there is more than one way of doing this we assume that one has been chosen.

Our goal is to find operators $\Phi(f)$ [we now drop the hats] such that

$$\Phi(Pf) = 0$$

for all $f \in C_0^\infty(M)$ and so that the time zero fields

$$\varphi(\underline{x}) = \Phi(0, \underline{x}) \qquad \pi(\underline{x}) = \dot{\Phi}(0, \underline{x})$$

(we have also removed the density ρ_h from the π) obey the CCRs

$$[\varphi(f), \pi(g)] = i\langle \overline{f} \mid g\rangle \mathbb{1}, \qquad [\varphi(f), \varphi(g)] = [\pi(f), \pi(g)] = 0, \tag{6}$$

for $f, g \in C_0^\infty(\Sigma)$. Here, the inner product on the right-hand side is that of $L^2(\Sigma, d\mathrm{vol}_h)$ [the ρ_h we dropped from π turns up in the integration measure]. In fact, this situation is of broader interest and it is convenient to work with square-integrable functions with respect to other measures on Σ with K self-adjoint on the new Hilbert space and using CCRs with respect to the new inner product. We simply write $L^2(\Sigma)$ to cover any such possibility. The extra freedom permits us to consider some nonglobally hyperbolic spacetimes with boundaries [with any necessary boundary conditions built in to the specification of the domain of K], and also the study of more general static spacetimes, after a redefinition of the fields. For example, if $M = \mathbb{R} \times \Sigma$ with a static[8] metric $g_{\alpha\beta}$, we define an 'optical metric' $k_{ij} = -g_{00}^{-1/2} g_{ij}$ on Σ. For a suitable operator of the form

$$K = -\triangle_k + m^2 + \text{curvature terms},$$

which is self-adjoint on $L^2(\Sigma, \rho_k)$, solutions ϕ to $\ddot{\phi} = K\phi$ yield solutions $\widetilde{\phi} = g_{00}^{(2-n)/4}\phi$ to the original Klein–Gordon equation on (M, g). We apply the CCRs using the inner product of $L^2(\Sigma, \rho_k)$.

Returning to the construction of the quantum field, first suppose that K has a spectrum consisting purely of strictly positive eigenvalues and that there is a corresponding basis of eigenvectors $\psi_j \in L^2(\Sigma) \cap C^\infty(\Sigma)$ for j in some index set $\mathscr{J}$. We may write

$$K\psi_j = \omega_j^2 \psi_j$$

for suitable ω_j that can be assumed positive. If Σ is compact these are comparatively mild assumptions, although it may be necessary to take m large enough to guarantee positivity of the spectrum.

To analyse the equation $\Phi(Pf) = 0$, we begin with f of the form $f(t, \underline{x}) = T(t)\psi_j(\underline{x})$, which gives us

$$\Phi((\ddot{T} + \omega_j^2 T)\psi_j) = 0$$

for all smooth compactly supported T. The general solution is

$$\Phi(T \otimes \psi_j) = \frac{1}{\sqrt{2\omega_j}} \int dt\, T(t) e^{-i\omega_j t} b_j + \frac{1}{\sqrt{2\omega_j}} \int dt\, T(t) e^{i\omega_j t} a_j^*$$

for operators a_j and b_j. The factors and adjoints here have been inserted for later convenience. In terms of the time-zero fields we have

$$\begin{aligned} b_j &= \frac{1}{\sqrt{2\omega_j}} \left(\omega_j \varphi(\psi_j) + i\pi(\psi_j)\right) \\ a_j^* &= \frac{1}{\sqrt{2\omega_j}} \left(\omega_j \varphi(\psi_j) - i\pi(\psi_j)\right). \end{aligned}$$

Noting that K commutes with complex conjugation, i.e., $K\overline{f} = \overline{Kf}$, we see that ψ_j and $\overline{\psi_j}$ must both be eigenvectors with the same eigenvalue. We may therefore assume that the basis has been chosen so that the set of basis vectors is invariant under complex

[8] $\partial g_{\alpha\beta}/\partial t = 0$, $g_{0i} = 0$.

conjugation, i.e., each ψ_j is either real, or $\overline{\psi_j} = \psi_{j'}$ for some other $j' \in \mathscr{J}$ (in which case ψ_j and $\overline{\psi_j}$ are orthogonal). A convenient way of expressing this is to write

$$\overline{\psi_j(\underline{x})} = \psi_{\overline{j}}(\underline{x}),$$

thinking of $j \mapsto \overline{j}$ as an involution on $\mathscr{J}$ (we're not thinking of this as complex conjugation); of course $\overline{j} = j$ if ψ_j is real-valued. This assumption has the advantage that $b_j = a_{\overline{j}}$ and the CCRs (6) simplify to

$$[a_j, a_k] = 0 \qquad [a_j, a_k^*] = \delta_{jk}\mathbb{1} \qquad j, k \in \mathscr{J}. \tag{7}$$

A general $\Phi(T \otimes S)$ ($T \in C_0^\infty(\mathbb{R})$, $S \in C_0^\infty(\Sigma)$) may be expressed by linearity and the basis property as

$$\Phi(T \otimes S) = \int dt\, T(t) \sum_{j \in \mathscr{J}} \frac{1}{2\sqrt{\omega_j}} \left(\langle \psi_j \mid S \rangle e^{-i\omega_j t} a_{\overline{j}} + \langle \psi_j \mid S \rangle e^{i\omega_j t} a_j^* \right)$$

or, after relabelling one of the summands

$$\Phi(T \otimes S) = \int dt\, T(t) \sum_{j \in \mathscr{J}} \frac{1}{2\sqrt{\omega_j}} \left(\langle \overline{\psi_j} \mid S \rangle e^{-i\omega_j t} a_j + \langle \psi_j \mid S \rangle e^{i\omega_j t} a_j^* \right).$$

We may summarise the construction above in as a recipe for constructing a Klein–Gordon QFT, applicable under the assumptions of this section.

Recipe: Choose an orthonormal basis $\{\psi_j\}$ for $L^2(\Sigma)$ of eigenfunctions for K, so that the set of basis vectors is closed under complex conjugation. Then quantum field may be represented [in unsmeared form] as

$$\Phi(t, \underline{x}) = \sum_{j \in \mathscr{J}} \frac{1}{\sqrt{2\omega_j}} \left(e^{-i\omega_j t} \psi_j(\underline{x}) a_j + e^{i\omega_j t} \overline{\psi_j(\underline{x})} a_j^* \right) \tag{8}$$

on a Hilbert space carrying a representation of the canonical commutation relations (7). We will describe how this Hilbert space can be constructed in the next subsection.

Note that the functions $u_j(t, \underline{x}) = (2\omega_j)^{-1/2} e^{-i\omega_j t} \psi_j(\underline{x})$ are all Klein–Gordon solutions; owing to the choice $\omega_j > 0$ they are called **positive frequency solutions**. Unlike the solutions in Sol, they do not generally have compact support on Cauchy surfaces (unless Σ is compact!). Nonetheless, Sol is dense in the complex Hilbert space with the u_j and the $\overline{u_j}$.

Example 1 Suppose Σ is a 3-torus, with a flat metric

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2$$

and periodicity a in each x_i. Then $K = -\triangle + m^2$ and we seek eigenfunctions on $[0, a] \times [0, a] \times [0, a]$ with appropriate periodic boundary conditions. This gives $\mathscr{J} = [(2\pi/a)\mathbb{Z}]^3$, and we have

$$\psi_{\boldsymbol{k}}(\boldsymbol{x}) = a^{-3/2} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$$

with

$$\omega_{\boldsymbol{k}} = \sqrt{\|\boldsymbol{k}\|^2 + m^2}.$$

Note that $\omega_0 = m$, so we will not obtain a ground state if $m = 0$. If $m > 0$, we have the field in the form

$$\Phi(t, \underline{x}) = \sum_{a\boldsymbol{k}/(2\pi)\in\mathbb{Z}^3} \frac{1}{\sqrt{2\omega_{\boldsymbol{k}} a^3}} \left(a_{\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t + i\boldsymbol{k}\cdot\boldsymbol{x}} + a_{\boldsymbol{k}}^* e^{i\omega_{\boldsymbol{k}} t - i\boldsymbol{k}\cdot\boldsymbol{x}} \right),$$

which is of course what is often meant by quantization in a box: it's really QFT in a spacetime with toroidal spatial section.

Example 2 The Einstein Universe is $\mathbb{R} \times S^3$ with metric

$$ds^2 = dt^2 - a^2(d\chi^2 + \sin^2\chi d\Omega^2),$$

where $d\Omega^2$ is the metric of the round unit 2-sphere, and $0 \le \chi \le \pi$. It has $R = 6/a^2$ and therefore solves the vacuum Einstein equations with cosmological constant $\Lambda = 3/(2a^2)$. The operator K is

$$K = -\frac{1}{a^2 \sin^2\chi} \left\{ \partial_\chi \sin^2 \partial_\chi + \triangle_{S^2} \right\} + m^2 + \frac{6\xi}{a^2},$$

and its eigenfunctions are conveniently obtained by separating into $F_\ell(\chi) Y_{\ell m}$, whereupon the F_ℓ must satisfy

$$-\frac{1}{\sin^2\chi} \partial_\chi \sin^2\chi \partial_\chi F_\ell + \frac{\ell(\ell+1)F_\ell}{\sin^2\chi} = ((\omega a)^2 - (ma)^2 - 6\xi) F_\ell$$

subject to Dirichlet boundary conditions at $\chi = 0, \pi$. For $\ell = 0$, it is not hard to see that we have a sequence of eigenfunctions of the form $(\sin k\chi)/\sin\chi$, $(k \in \mathbb{N})$ corresponding to

$$\omega_k^2 = \frac{k^2 - 1 + 6\xi}{a^2} + m^2.$$

As $\omega_1 \le 0$ if $\xi \le -(ma)^2/6$, there there is no ground state in this parameter range.

Many other examples may be found in [6].

4.2 Hilbert space constructions

The recipe given above requires the existence of a Hilbert space on which the canonical commutation relations (7) are represented. For quantum mechanical systems with finitely many degrees of freedom this is a straightforward matter. To take the simplest case of all, the CCR algebra $[a, a^*] = \mathbb{1}$ is represented on the domain of terminating sequences in $\ell^2(\mathbb{N}_0)$ by the weighted left-shift a and its adjoint, with actions

$$\begin{aligned} ae_n &= \sqrt{n} e_{n-1} \\ a^* e_n &= \sqrt{n+1} e_{n+1}, \end{aligned}$$

on the standard basis vectors e_n (by convention $e_{-1} = 0$). Moreover, any other Hilbert space representation (obeying certain conditions[9]) of this CCR algebra is unitarily equivalent to a direct sum of copies of the basic one just given. (This is the infinitesimal version of Stone–von Neumann uniqueness theorem; see, e.g., Example 1.3 in [1] for discussion and references.)

By contrast, a CCR algebra with infinitely many generators admits Hilbert space representations that cannot be placed in unitary equivalence. In this subsection we show how to construct a suitable representation for our recipe in two, equivalent, ways and also indicate how unitarily inequivalent representations may be constructed.

4.2.1 Tensor product construction

Our first construction simply forms a tensor product of $\mathscr{J}$-fold many copies of the basic representation described above. This is done as follows:

- Let S be the (countable) set of all sequences $(s_j)_{j\in\mathscr{J}}$ in $\mathbb{N}_0$ differing from 0 in only finitely many places.
- Construct a separable Hilbert space $\mathscr{H}$ with an orthonormal basis labelled by S, denoting the basis vector labelled by $(s_j)_{j\in\mathscr{J}}$ as
$$\bigotimes_{j\in\mathscr{J}} e_{s_j}.$$
We interpret the s_j as occupation numbers in mode j; a more standard physics notation would be $|s_{j_1}; s_{j_2}; \cdots\rangle$. By passing to bases, we immediately obtain a definition for any tensor product
$$\bigotimes_{j\in\mathscr{J}} \chi_j$$
in which each $\chi_j \in \ell^2(\mathbb{N}_0)$ and all but finitely many of the χ_j are e_0. The finite linear combinations of such vectors form a dense subspace of $\mathscr{H}$ denoted F.
- Let a_j act as a in the j'th factor, i.e.,
$$a_j\left(\bigotimes_{k\in\mathscr{J}} e_{s_k}\right) = \sqrt{s_j}\left(\bigotimes_{k\in\mathscr{J}} e_{s_k-\delta_{jk}}\right)$$
and
$$a_j^*\left(\bigotimes_{k\in\mathscr{J}} e_{s_k}\right) = \sqrt{s_j+1}\left(\bigotimes_{k\in\mathscr{J}} e_{s_k+\delta_{jk}}\right)$$

[9] For example, this applies to representations π in which $\pi(a)$ and $\pi(a^*)$ are defined on a common, dense, invariant domain $\mathscr{D}$ in the Hilbert space so that: (a) $\pi(a^*) = \pi(a)^*|_{\mathscr{D}}$; (b) $[\pi(a), \pi(a^*)]\psi = \psi$ for all $\psi \in \mathscr{D}$; and (c) $\pi(a^*)\pi(a)$ is essentially self-adjoint on $\mathscr{D}$.

and extend linearly to operators on F. The distinguished vector corresponding to the identically zero sequence in S

$$\Omega = \bigotimes_{j\in\mathscr{J}} e_0$$

is the vacuum vector and is annihilated by all of the a_j.

Other definitions of the infinite tensor product are possible. For example, instead of S, we could have chosen the set S' of sequences which are equal to 1 except in finitely many places. This yields a representation of the CCRs which is unitarily inequivalent to the one based on S. To see this, suppose they are unitarily equivalent under some U, and consider inner products of $U\Omega$ with the basis vectors of the new space, each of which [by definition of S'] is the image of some $\tilde{a}_j^*$ on some other basis vector. But each $\tilde{a}_j$ must annihilate $U\Omega$ by the supposed equivalence; thus $U\Omega$ is orthogonal to a basis for the space and we obtain a contradiction.

Of course there are very many variations on this theme. The representation we constructed using S is the only one that admits a total number operator,

$$N = \sum_{j\in\mathscr{J}} a_j^* a_j.$$

There is also a definition of the infinite tensor product, due to von Neumann, which includes all such possibilities in a single inseparable Hilbert space, but this is of less relevance to QFT.

4.2.2 Fock space

Our construction of the quantum field Φ and the Hilbert space $\mathscr{H}$ appears to depend on the particular basis ψ_j. We can remove this apparent dependence by constructing a unitarily equivalent Hilbert space representation of the CCRs: namely the bosonic Fock space over $L^2(\Sigma)$. This is given as a direct sum

$$\mathscr{F}_s(L^2(\Sigma)) = \bigoplus_{n=0}^{\infty} L^2(\Sigma)^{\odot n},$$

where $L^2(\Sigma)^{\odot n}$ denotes the symmetric n'th tensor power of $L^2(\Sigma)$, with the convention that this is $\mathbb{C}$ for $n = 0$. Explicitly,

$$L^2(\Sigma)^{\odot n} = \{F \in L^2(\Sigma^{\times n}) : F \text{ is symmetric in its arguments a.e.}\}.$$

Any element $\Psi \in \mathscr{F}$ can be represented as a sequence $\Psi^{(n)} \in L^2(\Sigma)^{\odot n}$, and we define operators $a(f)$, $a^*(f)$ for $f \in L^2(\Sigma)$ by

$$(a(f)\Psi)^{(n)}(\underline{x}_1, \ldots, \underline{x}_n) = \sqrt{n+1} \int_\Sigma d\mathrm{vol}_h(\underline{x}) \overline{f(\underline{x})} \Psi^{(n+1)}(\underline{x}, \underline{x}_1, \ldots, \underline{x}_n)$$

(warning: $a(f)$ is *anti*linear in f; some authors have $a(f)$ linear in f) and

$$(a^*(f)\Psi)^{(n)}(\underline{x}_1,\ldots,\underline{x}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n f(\underline{x}_i)\Psi^{(n-1)}(\underline{x}_1,\ldots,\widehat{\underline{x}_i},\ldots\underline{x}_n),$$

where the hat denotes an omitted variable. These definitions make sense on $\Psi \in F_0$, the states for which all but finitely many of the n-particle wavefunctions vanish. These annihilation and creation operators obey the commutation relations

$$[a(f), a(g)] = [a^*(f), a^*(g)] = 0, \qquad [a(f), a^*(g)] = \langle f \mid g\rangle \mathbb{1} \qquad f, g \in L^2(\Sigma)$$

and we also have $a^*(f) = a(f)^*$ for all $f \in L^2(\Sigma)$.

The connection with our previous recipe is

$$a_j = a(\psi_j) \qquad a_j^* = a^*(\psi_j)$$

and indeed $\mathscr{F}_s(L^2(\Sigma))$ and $\mathscr{H}$ may be identified in this way (more properly we ought to write a unitary isomorphism, but we suppress this), with the vacuum Ω being represented by $\Omega^{(0)} = 1$, $\Omega^{(n)} = 0$ for $n \geq 1$. The interpretation of $a^*(f)$ is that it creates a particle in a wavepacket, while a_j^* creates particle in a pure mode.

On $F_0 \subset \mathscr{F}_s(L^2(\Sigma))$, the field (or rather its smearings on surfaces of constant t) may now be represented without reference to a basis, as

$$\int_\Sigma d\mathrm{vol}_h(\underline{x})\Phi(t,\underline{x})f(\underline{x}) = \frac{1}{\sqrt{2}}\left(a(e^{iK^{1/2}t}K^{-1/4}\overline{f}) + a^*(e^{iK^{1/2}t}K^{-1/4}f)\right) \tag{9}$$

provided $f \in D(K^{-1/4})$. To see the connection with the previous recipe, note that (once (8) is spatially smeared) the two expressions agree for $f = \psi_j$ ($F(K)\psi_j = F(\omega_j^2)\psi_j$ by functional calculus) and hence for any finite linear combination of them. The advantage of (9) is that it is more easily generalized and permits us to drop many of the restrictions we placed on K, in particular, the requirement that it should have discrete spectrum. All that is required is that the expression $e^{iK^{1/2}t}K^{-1/4}f$ to exist as an element of $L^2(\Sigma)$ for every $f \in C_0^\infty(\Sigma)$ to define spatially smeared fields using (9). We may summarise this as follows.

Gourmet recipe: Let K be self-adjoint and positive[10] on (a dense domain in) $L^2(\Sigma)$ and suppose that $C_0^\infty(\Sigma) \subset D(K^{-1/4})$. Then the spatially smeared fields may be represented on the domain F_0 in Fock space $\mathscr{F}_s(L^2(\Sigma))$ by (9).

The usual Minkowski vacuum may be described in this way except in the case $n = 2$, $m = 0$, where the domain condition fails due to an infrared divergence. The conclusion is that there is no vacuum state on the algebra of fields in this case. One response is to work instead with the algebra of field derivatives; others are described in [34].

[10] In particular, the spectrum of K can include 0.

4.3 Ground states and thermal states

The Hilbert space construction presented above is known as a **ground state representation** because the vacuum vector Ω is a state of lowest eigenvalue for the Hamiltonian

$$H = \sum_{j \in \mathscr{J}} \omega_j a_j^* a_j,$$

which generates the time evolution in the sense that

$$e^{iH\tau}\Phi(t,\underline{x})e^{-iH\tau} = \Phi(t+\tau,\underline{x}).$$

Indeed, $H\Omega = 0$. In the Fock space picture we have

$$H = d\Gamma(\sqrt{K}) := 0 \oplus \sqrt{K} \oplus (\sqrt{K} \otimes \mathbb{1} + \mathbb{1} \otimes \sqrt{K}) \oplus \cdots$$

and

$$e^{iH\tau} = \Gamma(e^{i\sqrt{K}\tau}) := 1 \oplus e^{i\sqrt{K}\tau} \oplus \left(e^{i\sqrt{K}\tau} \otimes e^{i\sqrt{K}\tau}\right) \oplus \cdots;$$

accordingly, we refer to $\sqrt{K}$ as the **one-particle Hamiltonian** and H as its **second quantization**.

As a small digression, Hilbert spaces form a category Hilb with bounded linear maps as morphisms: the operation of forming a Fock space over a given Hilbert space is a functor from Hilb to itself acting on objects and morphisms by

$$\mathscr{H} \mapsto \mathscr{F}_s(\mathscr{H}) \qquad B(\mathscr{H},\mathscr{K}) \ni T \mapsto \Gamma(T) \in B(\mathscr{F}_s(\mathscr{H}),\mathscr{F}_s(\mathscr{K})),$$

which also preserves isomorphisms, i.e., if $U : \mathscr{H} \to \mathscr{K}$ is unitary, then so is $\Gamma(U)$. This is the origin of Nelson's phrase 'First quantization is a mystery; second quantization is a functor'.

Now let us consider the product of two fields. We calculate

$$\Phi(x)\Phi(x') = \sum_{j,j' \in \mathscr{J}} \frac{1}{2\sqrt{\omega_j \omega_{j'}}} \Big(a_j a_{j'}^* \psi_j(\underline{x})\overline{\psi_{j'}(\underline{x}')} e^{-i\omega_j t + i\omega_{j'} t'} + a_j^* a_{j'} \overline{\psi_j(\underline{x})}\psi_{j'}(\underline{x}') e^{-i\omega_j t + i\omega_{j'} t'}$$
$$+ \text{terms in } a_j^* a_{j'}^* \text{ and } a_j a_{j'} \Big),$$

so the two-point function of the ground state $G(x,x') = \langle \Omega \mid \Phi(x)\Phi(x')\Omega \rangle$ is

$$G(t,\underline{x};t',\underline{x}') = \sum_{j \in \mathscr{J}} \frac{1}{2\omega_j} \psi_j(\underline{x})\overline{\psi_j(\underline{x}')} e^{-i\omega_j(t-t')}.$$

Note that only contributions of 'positive frequency' with respect to t appear, i.e., $e^{-i\omega_j t}$. (We have chosen our nonstandard notion of Fourier transformation to make this positive frequency in that sense as well.)

Remarks:

1. Assuming that everything has gone according to plan, the advanced-minus-retarded fundamental solution should be given in terms of the antisymmetric part of G:
$$E(t,\underline{x};t',\underline{x}') = \sum_{j\in\mathscr{J}} \frac{1}{2i\omega_j}\left[\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}e^{-i\omega_j(t-t')} - \psi_j(\underline{x}')\overline{\psi_j(\underline{x})}e^{i\omega_j(t-t')}\right]. \qquad (10)$$
It is a useful exercise to check this directly.

2. We may usefully think of $G(t,\underline{x};t',\underline{x}')$ as the integral kernel of the operator
$$(2\sqrt{K})^{-1}e^{-i\sqrt{K}(t-t')}$$
on $L^2(\Sigma)$. In this form, we may also address cases where the spectrum is not discrete. As already mentioned, we need to have $C_0^\infty(M) \subset D(K^{-1/4})$ to ensure that we really can smear our field against arbitrary test functions.

As well as the ground state, we shall also be interested in thermal equilibrium states. In the simple case where K has discrete spectrum we can make sense of the Gibbs state at temperature T, with expectation values
$$\langle A\rangle_\beta = \frac{\mathrm{Tr}\, e^{-\beta H}A}{\mathrm{Tr}\, e^{-\beta H}},$$
where $\beta = (kT)^{-1}$, provided the number of eigenvalues of K below ω^2 (counting multiplicity) grows polynomialy with ω, for then $e^{-\beta H}$ is trace-class. This will be true, for example, where K is formed from the Laplacian of a compact Riemannian manifold.

For a single harmonic oscillator we can easily show that
$$\frac{\mathrm{Tr}\, e^{-\beta\omega a^*a}aa^*}{\mathrm{Tr}\, e^{-\beta\omega a^*a}} = \frac{1}{1-e^{-\beta\omega}}$$
and
$$\frac{\mathrm{Tr}\, e^{-\beta\omega a^*a}a^*a}{\mathrm{Tr}\, e^{-\beta\omega a^*a}} = \frac{1}{e^{\beta\omega}-1},$$
while the operators a^2 and $(a^*)^2$ have vanishing expectation values. Thus the two-point function is
$$G_\beta(t,\underline{x};t',\underline{x}') = \sum_{j\in\mathscr{J}} \frac{1}{2\omega_j}\left(\frac{\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}e^{-i\omega_j(t-t')}}{1-e^{-\beta\omega_j}} + \frac{\overline{\psi_j(\underline{x})}\psi_j(\underline{x}')e^{i\omega_j(t-t')}}{e^{\beta\omega_j}-1}\right)$$
or, relabelling the second sum,
$$G_\beta(t,\underline{x};t',\underline{x}') = \sum_{j\in\mathscr{J}} \frac{1}{2\omega_j}\frac{\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}}{1-e^{-\beta\omega_j}}\left(e^{-i\omega_j(t-t')} + e^{-\beta\omega_j}e^{i\omega_j(t-t')}\right),$$
which now has an admixture of positive and negative frequencies with respect to t. Note, however, that the negative frequency part is exponentially suppressed.

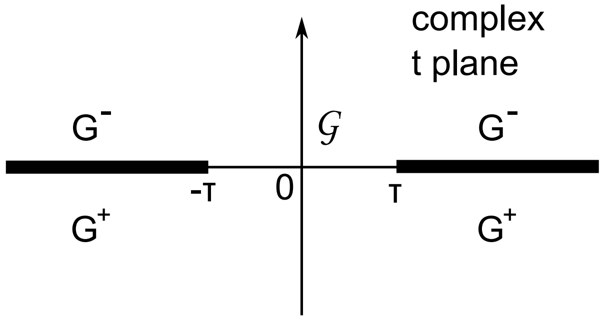


Figure 3: The cut plane for the vacuum two-point function

This two-point function is also the integral kernel of an operator, namely:

$$(2\sqrt{K})^{-1}(1-e^{-\beta\sqrt{K}})^{-1}\left(e^{-i\sqrt{K}(t-t')}+e^{-\beta\sqrt{K}}e^{i\sqrt{K}(t-t')}\right)$$

permitting us to generalize to K with nondiscrete spectrum: the criterion for the existence of thermal equilibrium states is then $C_0^\infty(\Sigma)\subset D(K^{-1/2})$, which is stronger than the ground state condition. For example, in Minkowski space these states do not exist in $n=2,3$ if $m=0$. In general, the states constructed are no longer Gibbs states, but obey the more general KMS condition described in Prof Bach's lectures.

Of course, the antisymmetric part of the 2-point function should still be a multiple of E – again, this can easily be checked.

4.4 Analytic structure and the Euclidean Green function

It is convenient to denote

$$G^\pm(t;\underline{x},\underline{x}')=\sum_{j\in\mathscr{J}}\frac{e^{\mp i\omega_j t}}{2\omega_j}\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}$$

so that

$$G^+(t;\underline{x},\underline{x}')=\langle\Omega\mid\Phi(t,\underline{x})\Phi(0,\underline{x}')\Omega\rangle\quad\text{and}\quad G^-(t;\underline{x},\underline{x}')=\langle\Omega\mid\Phi(0,\underline{x}')\Phi(t,\underline{x})\Omega\rangle$$

If $\underline{x}\neq\underline{x}'$ and $|t|$ is sufficiently small, these functions should coincide [commutation at spacelike separation]. Moreover, $G^\pm$ have analytic continuations in the lower$(+)$/upper$(-)$ complex half-planes in t, which evidently match across some interval $[-\tau,\tau]$ of the real t-axis. Accordingly, we deduce the existence of a holomorphic function $\mathcal{G}$ on the cut plane $\mathbb{C}\backslash((-\infty,-\tau)\cup(\tau,\infty))$ so that

$$G^\pm(t;\underline{x},\underline{x}')=\lim_{\epsilon\to 0^+}\mathcal{G}(t\mp i\epsilon;\underline{x},\underline{x}')$$

for $t\in\mathbb{R}$ – see Fig. 3.

On the imaginary axis we have

$$\mathcal{G}(is;\underline{x},\underline{x}')=\sum_{j\in\mathscr{J}}\frac{e^{-\omega_j|s|}}{2\omega_j}\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}$$

for $s \neq 0$. Mode-by-mode, this is not even smooth in s, let alone holomorphic, but the series sums to a holomorphic limit nonetheless. It is instructive to examine the coefficient of $\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}$ in the series for $\mathcal{G}(i(s-s');\underline{x},\underline{x}')$, namely

$$\frac{e^{-\omega_j|s-s'|}}{2\omega_j}.$$

Observing that this is the Green function for the operator $-\partial^2/\partial s^2+\omega_j^2$, Hilbert space magic allows us to conclude that

$$G_E(s,\underline{x};s',\underline{x}') = \mathcal{G}(i(s-s');\underline{x},\underline{x}')$$

is the integral kernel of the operator

$$\left(-\frac{\partial^2}{\partial s^2}+K\right)^{-1}$$

on $L^2(\mathbb{R}\times\Sigma)$, i.e., the so-called **Euclidean Green function**. Note that the operator $-\partial^2/\partial s^2+K$ is elliptic, and therefore has a unique Green function, in contrast to the hyperbolic Klein–Gordon operator.

Example: Consider the massive scalar field in n-dimensional Minkowski space, for which the Euclidean Green function is

$$G_E(x,x') = \int \frac{d^nk}{(2\pi)^n}\frac{e^{ik\cdot(x-x')}}{|k|_E^2+m^2},$$

where $|\cdot|_E$ is the Euclidean norm. Straightforward manipulations give $G_E(x,x') = F(|x-x'|_E)$ where

$$F(z) = \frac{\pi\mathrm{Area}(S^{n-2})}{(2\pi)^n}\int_m^\infty d\omega e^{-z\omega}(\omega^2-m^2)^{(n-3)/2}$$

is evidently holomorphic in $\mathrm{Re}\, z>0$. For example, in $n=4$,

$$F(z) = \frac{m}{4\pi^2 z}K_1(mz) = \frac{1}{4\pi^2z^2}+\frac{m^2}{8\pi^2}\log(mz)+O(1).$$

Thus, for fixed $\underline{x}\neq\underline{x}'$,

$$\mathcal{G}(is;\underline{x},\underline{x}') = F([s^2+|\underline{x}-\underline{x}'|^2]^{1/2})$$

is holomorphic on the cut s-plane with cuts extending from the branch points at $s = \pm i|\underline{x}-\underline{x}'|$ to infinity along the imaginary axis.

By analytic continuation, the vacuum 2-point function can now be expressed as the boundary value

$$\langle\Omega\mid\Phi(t,\underline{x})\Phi(t',\underline{x}')\Omega\rangle = \lim_{\epsilon\to0^+}F([-(t-t'-i\epsilon)^2+|\underline{x}-\underline{x}'|^2]^{1/2}).$$

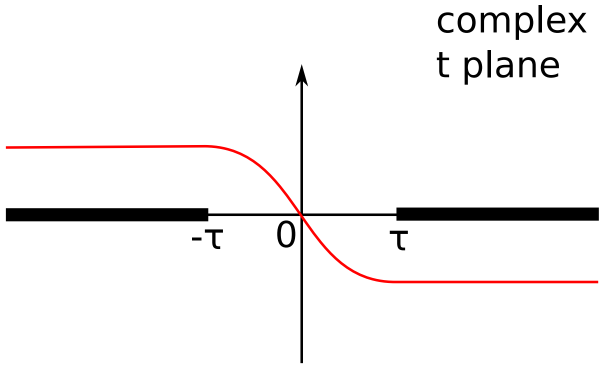


Figure 4: The Feynmann propagator is obtained as the boundary value of $\mathcal{G}$ taken as the red contour approaches the real axis.

Hence, for $n = 4$,

$$\langle \Omega \mid \Phi(t, \underline{x})\Phi(t', \underline{x}')\Omega \rangle = \frac{1}{4\pi^2\sigma_+} + \frac{m^2}{16\pi^2} \log m^2\sigma_+ + \dots,$$

where by $f(\sigma_+)$ we mean the distributional limit

$$f(\sigma_+) = \lim_{\epsilon \to 0^+} f(-(t - t' - i\epsilon)^2 + |\underline{x} - \underline{x}'|^2).$$

As we will see later this singular structure is universal in a certain sense. (Warnings: (i) σ is negative for timelike separation and positive for spacelike separation, in contrast to our signature convention; (ii) some authors use σ to denote a multiple of the squared geodesic separation.)

Remarks:

1. The branch points are responsible for singularities of $\langle \Omega \mid \Phi(x)\Phi(x')\Omega \rangle$ occuring precisely where x and x' are null related.

2. The cuts mean that analytic continuation needs to be preformed carefully. For example, as illustrated in Fig. 3, the two-point function G_+ (G_-) is the boundary value of $\mathcal{G}$ as t approaches the real-axis from below (above). However a rigid Wick rotation of the Euclidean Green function produces a different boundary value, namely the Feynmann propagator

$$G_F(t; \underline{x}, \underline{x}') = \lim_{\theta \to \pi/2^-} \mathcal{G}(ite^{i\theta}; \underline{x}, \underline{x}')$$

and because the boundary value is taken on contours that pass under one cut but over the other (see Fig. 4 for an equivalent contour), we have

$$G_F(t; \underline{x}, \underline{x}') = \begin{cases} G^+(t; \underline{x}, \underline{x}') & t > 0 \\ G^-(t; \underline{x}, \underline{x}') & t < 0. \end{cases}$$

This is all nicely explained in [34].

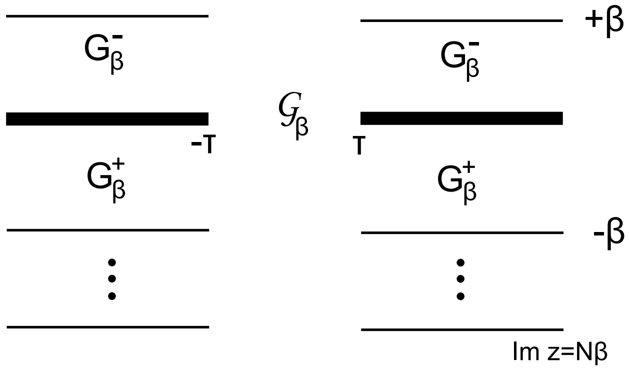


Figure 5: The cut plane for the thermal two-point function

Performing a similar analysis for the thermal equilibrium states, we set

$$G_\beta^\pm(t;\underline{x},\underline{x}') = \sum_{j\in\mathscr{J}} \frac{\psi_j(\underline{x})\overline{\psi_{j'}(\underline{x}')}}{2\omega_j(1-e^{-\beta\omega_j})} q_\pm(t),$$

where

$$q_\pm(z) = e^{\mp i\omega_j z} + e^{\pm i\omega_j(z\pm i\beta)}.$$

The formulae for $G_\beta^\pm$ define analytic functions in $-\beta < \operatorname{Im} z < 0$ $(+)$ or $0 < \operatorname{Im} z < \beta$ $(-)$; moreover these analytic functions coincide on an interval of the real t-axis if $\underline{x} \neq \underline{x}'$, thus forming a single holomorphic function $\mathcal{G}_\beta(z;\underline{x},\underline{x}')$ on $\{z \in \mathbb{C} : |\operatorname{Im} z| < \beta\}\setminus((-\infty,-\tau)\cup(\tau,\infty))$.

We also have the important identity $q_+(z-i\beta) = q_-(z)$ for any z, [consistent with the KMS condition] and this implies that $\mathcal{G}_\beta$ may be continued further using the identity

$$\mathcal{G}_\beta(z;\underline{x},\underline{x}') = \mathcal{G}_\beta(z+iN\beta;\underline{x},\underline{x}')$$

for all integers N – see Fig. 5.

Passing to the imaginary t axis, we have

$$\mathcal{G}_\beta(is;\underline{x},\underline{x}') = \sum_{j\in\mathscr{J}} \frac{\psi_j(\underline{x})\overline{\psi_j(\underline{x}')}}{2\omega_j(1-e^{-\beta\omega_j})} \left(e^{-\omega_j s} + e^{\omega_j(s-\beta)}\right)$$

for $0 < s < \beta$, periodically continued outside this interval. We claim that $\mathcal{G}_\beta(i(s-s');\underline{x},\underline{x}')$ is the Euclidean Green function on $L^2(\mathbb{T}_\beta\times\Sigma)$, where $\mathbb{T}_\beta$ denotes a circle of periodicity β. This may be verified mode by mode by observing that

$$\frac{1}{2\omega_j(1-e^{-\beta\omega_j})}\left(e^{-\omega_j|s-s'|} + e^{\omega_j(|s-s'|-\beta)}\right)$$

is the Green function for $-\partial^2/\partial s^2 + \omega_j^2$ on $L^2(\mathbb{T}_\beta)$. This is what is meant by phrases such as 'finite temperature equates to periodicity in imaginary time'.

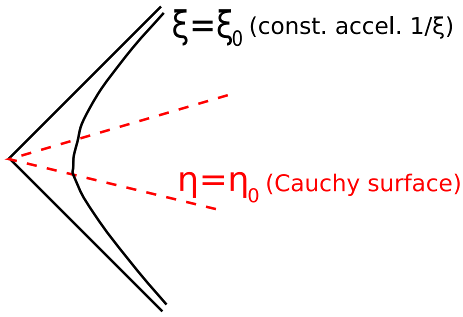


Figure 6: Diagram of Rindler space, showing lines of constant η and ζ.

4.5 The Unruh effect

We now come to one of the early surprises of QFT in CST. While attempting to understand Hawking's prediction of radiation from black holes [37], Unruh discovered an analogous effect: a uniformly accelerated detector in Minkowski space will detect a thermal spectrum of particles in the Minkowski vacuum [67]. For a recent review of the extensive literature on the Unruh effect, see [11]. Here, we show how the Unruh effect may be explained using the techniques of the previous section, again following [34].

We consider uniformly accelerated trajectories confined to the Rindler wedge $x > |t|$ of n-dimensional Minkowksi space, where $(t, x, \underline{x}_\perp)$ are the standard Minkowski coordinates. Replacing the coordinates t and x by η and ξ so that

$$t = \xi \sinh \eta \qquad x = \xi \cosh \eta,$$

the Rindler wedge is covered by the coordinate range $(\eta, \xi) \in \mathbb{R} \times (0, \infty)$, $\underline{x}_\perp \in \mathbb{R}^{n-2}$, and has metric

$$ds^2 = \xi^2 d\eta^2 - d\xi^2 - \delta_{ij} dx^i_\perp dx^j_\perp.$$

As shown in Fig. 6, lines of constant η are Cauchy surfaces, while a curve of constant ζ is a uniformly accelerated trajectory with proper acceleration ξ^{-1}. The acceleration horizon corresponds to the portions of the null lines $t = \pm x$ lying in $x > 0$.

With these coordinates, Rindler space fits into our general static (though not ultrastatic) framework, as the Klein–Gordon equation is

$$\frac{\partial^2 \phi}{\partial \eta^2} + K\phi = 0,$$

where

$$K = -(\xi^2 \partial_\xi^2 + \xi \partial_\xi + \triangle_\perp) + m^2 \xi^2$$

is self-adjoint on a dense domain in $L^2(\mathbb{R} \times (0, \infty) \times \mathbb{R}^{n-2}, \xi^{-1} d\xi d^{n-2} \underline{x}_\perp)$.

Accordingly, the thermal equilibria states with respect to the 'time' parameter η correspond to Euclidean Green functions on metrics

$$ds^2 = d\xi^2 + \xi^2 d\theta^2 + \delta_{ij} dx^i_\perp dx^j_\perp$$

(we discard an overall minus for simplicity) with $\theta \in (0, \beta)$ periodically identified. This corresponds to a space with a conical singularity at the origin except for the special case $\beta = 2\pi$, where it is just the metric of Euclidean 4-space in polar coordinates, i.e., $x^0 = \xi \cos\theta$, $x^1 = \xi \sin\theta$. Thus $\mathcal{G}^{\mathrm{R}}_{2\pi}$ is just $\mathcal{G}^{\mathrm{M}}$ in different coordinates; in fact,

$$\mathcal{G}^{\mathrm{R}}_{2\pi}(\eta - \eta'; \xi, \underline{x}_\perp; \xi', \underline{x}'_\perp) = \mathcal{G}^{\mathrm{M}}(\xi \sinh\eta - \xi' \sinh\eta'; \xi \cosh\eta - \xi' \cosh\eta', \underline{x}_\perp - \underline{x}'_\perp).$$

In other words, the restriction of the Minkowski vacuum state to the Rindler wedge coincides with the thermal equilibrium state at inverse 'temperature' 2π with respect to the Rindler 'time' coordinate η. Adjusting for the normalisation of $\partial/\partial\eta$, we find that the temperature is $1/(2\pi\xi)$, and diverges as the horizon is approached ($\xi \to 0^+$).

For a uniformly accelerated observer following a path of constant ξ, the parameter η would be a natural time parameter: such an observer would therefore regard the Minkowski vacuum as 'hot' (although an acceleration of $10^{19}\mathrm{ms}^{-2}$ is needed for a temperature of 1K!) We will return to the issue of particle detectors at the end of this subsection.

The Euclideanisation trick provides a number of interesting states in other spacetimes. For example, de Sitter spacetime is a vacuum solution ($T_{ab} = 0$) to Einstein's equations with cosmological constant $\Lambda = 3/\alpha^2$, and may be regarded as the hyperboloid

$$T^2 - S^2 - X^2 - Y^2 - Z^2 = -\alpha^2,$$

where (T, S, X, Y, Z) are inertial coordinates on 5-dimensional Minkowski space. In this form it is clear that the Euclidean form of de Sitter is nothing but the round 4-sphere. Taking the Euclidean vacuum in this case gives the **Gibbons–Hawking state** on de Sitter [35], which is invariant under the de Sitter symmetries and produces a thermal spectrum of particles with respect to the proper time along any timelike geodesic.

A similar argument can be made in the case of the four-dimensional Schwarzschild black hole, with metric

$$ds^2 = (1 - 2M/r)dt^2 - (1 - 2M/r)^{-1}dr^2 - r^2 d\Omega^2,$$

where $d\Omega^2$ is the usual metric on S^2. The metric component g_{rr} diverges at $r = 2M$, but this is a defect of the coordinates rather than an actual singularity: $r = 2M$ is the event horizon of the black hole (note that $\partial/\partial t$ becomes null there). A formal analytic continuation of $\tau = it$ gives the metric [again discarding an overall sign]

$$ds_E^2 = (1 - 2M/r)d\tau^2 + (1 - 2M/r)^{-1}dr^2 + r^2 d\Omega^2,$$

which possesses a conical defect at $r = 2M$ unless τ is chosen to be periodic with period $8\pi M$ [expanding $r = 2M + \epsilon$ we have

$$ds_E^2 = (1 - 2M/r)^{-1}\left(\frac{\epsilon^2}{(4M)^2}d\tau^2 + d\epsilon^2\right) + r^2 d\Omega^2$$

so we need $\tau/(4M)$ to have period 2π.] The upshot is the **Hartle–Hawking state** defined on Kruskal space – an extension of Schwarzschild – which restricts to Schwarzschild as a thermal equilibrium state at $\beta = 8\pi M$ with respect to the t-coordinate. The analogy with the Unruh effect suggests that this state has a special status, and that black holes like their surroundings to be in thermal equilibrium at the **Hawking temperature**

$$T_H = \frac{1}{8\pi k M},$$

as seen by 'static observers at infinity' (whose proper time coincides with t). This is supported by the rigorous results of Kay and Wald [49], who proved that any stationary nonsingular state on Kruskal would have to restrict to Schwarzschild as a Hawking-temperature thermal state. Nonetheless, the question of whether this state exists [i.e., whether the Hartle-Hawking state is indeed nonsingular etc] has not yet been settled. Of course, these results do not directly address Hawking's original model of black hole radiation, which concerns a collapsing body.

The key lesson to be gained from this subsection is that the Fock space quantization we have developed is far from being covariant. The constructions of ground and thermal states depend critically on the coordinate used as 'time', and a different choice of coordinate will generally result in a different state. Here we have a key difference with ordinary QFT, because in a curved spacetime theory we expect invariance under general coordinate transformations. So the Fock space cannot be an invariant object – which means that the concept of particle is also noninvariant. QFT in CST is truly a theory of fields, not of particles.

Further support for this view comes from particle detector models which couple the field to an auxiliary quantum system. In first order perturbation theory one finds that a uniformly accelerated detector in Minkowski space detects a thermal spectrum of particles in the vacuum state, while an inertial detector is not triggered. For a discussion of the perturbative approach, taking care of the switching of the detector and general trajectories, see [52]. A fully rigorous analysis of the detector system has been made in terms of a 'return to equilibrium' problem [14], justifying the perturbative results in a specified regime of weak coupling between the detector and the field. However, the detector response may be rather different under other circumstances [51]: in an exact treatment, even a detector at rest in the Minkowski vacuum becomes entangled with the field, leading to a mixed detector state when the field is traced out.

4.6 n-point functions for the Fock vacuum

So far, we have restricted attention to the two-point functions of ground and thermal states. We now briefly discuss n-point functions such as

$$\langle \Omega \mid \Phi(f_1) \cdots \Phi(f_n) \Omega \rangle$$

and discuss their relation to the two-point function. For simplicity, we return to the ground state Ω in the case where K has purely discrete spectrum. Then the n-point function can be expanded as a sum of terms each involving a factor of the form

$$\langle \Omega \mid a_1^{\flat_1} \cdots a_n^{\flat_n} \Omega \rangle, \tag{11}$$

where each $\flat$ denotes either the presence or absence of an adjoint. It is easy to see that not all of these terms can contribute: for example, any term with $a_n^{\flat_n} = a_n$ or $a_1^{\flat_1} = a_1^*$ must vanish. More generally, we can evaluate any such term by commuting all the annihilation operators to the right. Each time an annihilation operator a_p passes to the right of a creation operator a_q^*, their commutator produces another term of the form (11) but with a_p and a_q^* deleted and multiplied by a factor of $\langle \Omega \mid a_p a_q^* \Omega \rangle$. Proceding recursively, every term is reduced to a sum of products of such factors.

By a bit of brooding, one realises that this process may be represented using directed graphs. A **quasifree graph** on n-vertices labelled $1, \dots, n$ is a directed graph such that

- each vertex is met by precisely one edge, so each vertex is either a source or a target;
- for each edge e, we have $t(e) > s(e)$, where $t(e)$ is the target vertex of edge e and $s(e)$ is its source vertex.

We denote the set of all quasigraphs on n-vertices by $\mathcal{G}_n$; evidently $\mathcal{G}_n$ is empty if n is odd, while

$$\mathrm{Card}(\mathcal{G}_{2n}) = (2n-1)(2n-3)\dots 1 = \frac{(2n)!}{2^n n!}.$$

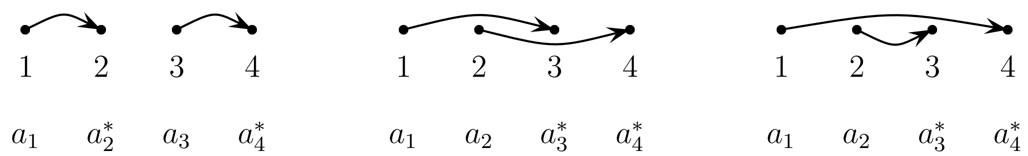
(There are $2n-1$ possible targets for the edge sourced at vertex 1, $2n-3$ possible targets for the edge with the next lowest source vertex, etc etc.)

It turns out that the inner product (11) is nonvanishing only when one or more quasifree graphs on n-vertices may be drawn so that vertex j is a target if and only if $\flat_j = *$. Moreover the value of the inner product (11) is

$$\langle \Omega \mid a_1^{\flat_1} \cdots a_n^{\flat_n} \Omega \rangle = \sum_G \prod_{e \in \mathrm{Edge}(G)} \langle \Omega \mid a_{s(e)} a_{t(e)}^* \Omega \rangle,$$

where the sum is taken over the quasifree graphs G corresponding to $\flat_1, \dots, \flat_n$.

Example: $\mathcal{G}_4$ consists of three graphs, corresponding to strings of annihilation and creation operators as shown. Note that the second and third graphs correspond to the *same* string of operators.



The first of these graphs corresponds to the identity

$$\langle \Omega \mid a_1 a_2^* a_3 a_4^* \Omega \rangle = \langle \Omega \mid a_1 a_2^* \Omega \rangle \langle \Omega \mid a_3 a_4^* \Omega \rangle,$$

while the second two graphs correspond to

$$\langle \Omega \mid a_1 a_2 a_3^* a_4^* \Omega \rangle = \langle \Omega \mid a_1 a_3^* \Omega \rangle \langle \Omega \mid a_2 a_4^* \Omega \rangle + \langle \Omega \mid a_1 a_4^* \Omega \rangle \langle \Omega \mid a_2 a_3^* \Omega \rangle .$$

The upshot of all this is that the n-point function of the Fock vacuum state Ω decomposes as a sum of products of its two-point functions

$$\langle \Omega \mid \Phi(f_1) \cdots \Phi(f_n) \Omega \rangle = \sum_{G \in \mathcal{G}_n} \prod_{e \in \mathrm{Edge}(G)} \langle \Omega \mid \Phi(F_{s(e)}) \Phi(F_{t(e)}) \Omega \rangle ,$$

and, in consequence, all its odd n-point functions vanish. Any state admitting this decomposition is said to be **quasifree**. (Some authors would permit a nonvanishing 1-point function in the definition of 'quasifree'.)

In particular, for any $f \in C_0^\infty(M; \mathbb{R})$ we have

$$\langle \Omega \mid \Phi(f)^{2n} \Omega \rangle = \frac{(2n)!}{2^n n!} \langle \Omega \mid \Phi(f)^2 \Omega \rangle^n ,$$

which entails (at least formally, although one can actually prove this rigorously)

$$\langle \Omega \mid e^{i\Phi(f)} \Omega \rangle = \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} \langle \Omega \mid \Phi(f)^{2n} \Omega \rangle = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \langle \Omega \mid \Phi(f)^2 \Omega \rangle^n = e^{-\|\Phi(f)\Omega\|^2/2}. \quad (12)$$

4.7 Nonstatic situations: particle production and the Hawking effect

It is reasonably clear how our recipe for quantization can be extended to general spacetimes: the field would be given in the form

$$\Phi(x) = \sum_{j \in \mathscr{J}} \left(u_j(x) a_j + \overline{u_j(x)} a_j^* \right) ,$$

where $\mathscr{J}$ is some index set and the u_j are Klein–Gordon solutions that, together with their complex conjugates, form a basis for a Hilbert space completion of Sol, and so that the commutation relation

$$[\Phi(x), \Phi(x')] = iE(x, x')\mathbb{1}$$

holds. This last is equivalent to

$$\sum_{j \in \mathscr{J}} \left(u_j(x) \overline{u_j(x')} - \overline{u_j(x)} u_j(x') \right) = iE(x, x'). \quad (13)$$

In many circumstances $\mathscr{J}$ would be a continuous index set and the sum is written simply for notational convenience.[11]

The choice of the u_j (or the space they span) in this construction is by no means unique. Suppose for example that $M = \mathbb{R} \times \Sigma$ and the metric obeys

$$ds^2 = \begin{cases} dt^2 - h_{ij}^{\mathrm{in}} dx^i dx^j & t < t_0 \\ dt^2 - h_{ij}^{\mathrm{out}} dx^i dx^j & t > t_1 . \end{cases}$$

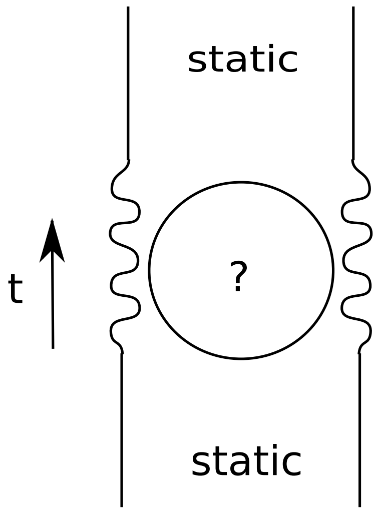


Figure 7: Schematic representation of the nonstatic spacetime example.

A schematic representation of this spacetime is given in Fig. 7.

Then there are two obvious candidates for a 'good' basis, namely $u_j^{\text{in/out}}$ labelled by possibly different index sets $\mathscr{J}^{\text{in/out}}$ and fixed by the conditions

$$\begin{aligned} u_j^{\text{in}}(t,\underline{x}) &= (2\omega_j^{\text{in}})^{-1/2} e^{-i\omega_j^{\text{in}} t}\psi_j^{\text{in}}(\underline{x}) & t < t_0 \\ u_j^{\text{out}}(t,\underline{x}) &= (2\omega_j^{\text{out}})^{-1/2} e^{-i\omega_j^{\text{out}} t}\psi_j^{\text{out}}(\underline{x}) & t > t_1, \end{aligned}$$

where the $\psi_j^{\text{in/out}}$ are eigenfunctions of the operators $K^{\text{in/out}}$ that would be associated with the in and out regions. Applying our usual recipe in each case gives two quantizations, with fields $\Phi^{\text{in/out}}$ defined on $\mathscr{F}^{\text{in/out}}$, based on vectors $\Omega^{\text{in/out}}$ that would represent vacua for observers at early/late times respectively. *A priori* it is not clear that there is a unitary equivalence between them, by which we mean a unitary $U : \mathscr{F}^{\text{in}} \to \mathscr{F}^{\text{out}}$ such that

$$U\Phi^{\text{in}}(f)U^{-1} = \Phi^{\text{out}}(f)$$

for all $f \in C_0^\infty(M)$; even if there is such a unitary, it is easily seen that $U\Omega^{\text{in}}$ will not in general be a scalar multiple of Ω^{out} – in other words the two quantizations assign inequivalent meanings to presence/absence of particles.

The reason for this is that a solution u_j^{in} will generally be a nontrivial linear combination of $u_{j'}^{\text{out}}$'s *and* $\overline{u_{j'}^{\text{out}}}$'s: in other words the u_j^{in} and $u_{j'}^{\text{out}}$ span distinct subspaces of possibly distinct completions of Sol. If

$$u_j^{\text{in}} = \sum_{j' \in \mathscr{J}^{\text{out}}} \left(\alpha_{jj'} u_{j'}^{\text{out}} + \beta_{jj'} \overline{u_{j'}^{\text{out}}} \right)$$

[11]There are more satisfactory basis-free descriptions (see, e.g., Wald's treatment [71]), but we proceed formally in order to capture the main idea.

then

$$\begin{aligned}\Phi^{\text{in}} &= \sum_{j\in\mathscr{J}^{\text{in}}}\sum_{j'\in\mathscr{J}^{\text{out}}}\left(\alpha_{jj'}u_{j'}^{\text{out}} + \beta_{jj'}\overline{u_{j'}^{\text{out}}}\right)a_j^{\text{in}} + \text{h.c.}\\ &= \sum_{j'\in\mathscr{J}^{\text{out}}}\sum_{j\in\mathscr{J}^{\text{in}}}\left(\alpha_{jj'}a_j^{\text{in}} + \overline{\beta_{jj'}}a_j^{\text{in}*}\right)u_{j'}^{\text{out}} + \text{h.c.}\end{aligned}$$

from which we may infer

$$U^{-1}a_{j'}^{\text{out}}U = \sum_{j\in\mathscr{J}^{\text{in}}}\left(\alpha_{jj'}a_j^{\text{in}} + \overline{\beta_{jj'}}a_j^{\text{in}*}\right)$$

and hence

$$\langle U\Omega^{\text{in}} \mid N^{\text{out}}U\Omega^{\text{in}}\rangle = \sum_{\substack{j\in\mathscr{J}^{\text{in}}\\ j'\in\mathscr{J}^{\text{out}}}}|\beta_{jj'}|^2,$$

which quantifies the extent of particle creation observed at late times if the state represents a vacuum at early times.

Remarks:

1. The state is not evolving, we are just expressing the initial vacuum in terms of the out-Fock space.

2. The expected number of particles created from the vacuum will be finite if $\beta_{jj'}$ obeys a Hilbert-Schmidt condition. It may be shown [71] that this is necessary and sufficient for the existence of the unitary U [provided that the u_j^{in} and $\overline{u_j^{\text{in}}}$ span the same completion of Sol as the $u_{j'}^{\text{out}}$ and $\overline{u_{j'}^{\text{in}}}$].

3. Both the 'in' and 'out' modes have to satisfy (13), which imposes constraints on the coefficients $\alpha_{jj'}$ and $\beta_{jj'}$; in fact:

$$\sum_{j'\in\mathscr{J}^{\text{out}}}\left(\alpha_{jj'}\overline{\alpha_{kj'}} - \beta_{jj'}\overline{\beta_{kj'}}\right) = \delta_{jk}$$

   and

$$\sum_{j'\in\mathscr{J}^{\text{out}}}\left(\alpha_{jj'}\beta_{kj'} - \beta_{jj'}\alpha_{kj'}\right) = 0.$$

   The $\alpha_{jj'}$ and $\beta_{jj'}$ are called **Bogolubov coefficients** and the transformation between the two representations is known as a **Bogolubov transformation**. A more satisfactory basis-independent account of the above theory can be found in Wald's book [71].

4. Rapid expansion in the early universe will have created particles and left an imprint on e.g., spectrum of density fluctuations.

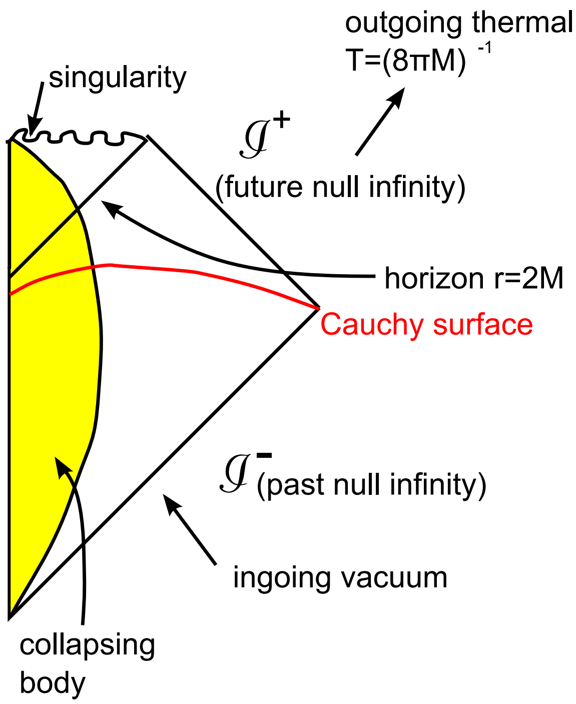


Figure 8: Conformal diagram illustrating the Hawking effect.

5. The Hawking effect [37] provides perhaps the most famous example of particle creation. Consider a star of mass M collapsing to form a Schwarzschild black hole, as shown on the conformal diagram Fig. 8. Before collapse begins we have a static situation; the same is true in the asymptotic far future. Then a state which is static and vacuum at early times will present a thermal spectrum of particles from the viewpoint of an observer far from the black hole at late times.

6. One might think that the 'correct' vacuum is neither Ω^{in} or Ω^{out}, but some other state which would treat the in and out regions democratically. A variety of schemes of this sort have been proposed, but they are necessarily global and/or restricted to particular subclasses of spacetimes, and often result in unphysical states. See, for example, the comments in sections 3 and 7 of [48]. The fundamental problem is that there is no *covariant* means of selecting a single state in each globally hyperbolic spacetime [9]. Recent methods that *do* yield physically acceptable states in certain types of spacetime are due to Olbermann [55] (Robertson-Walker models) and Moretti and collaborators [12, 53] (asymptotically flat globally hyperbolic spacetimes).

7. All this provides further evidence that the particle concept is not at all an invariant element of the theory. This posed the originators of QFT in CST many problems, because it is so in-grained a part of QFT. How many other notions would turn out to be observer-dependent? Gibbons and Hawking, in a 1977 paper, went as far as to say:

   > It would seem that one cannot, as some authors have attempted, construct a unique observer-independent renormalized energy-momentum

tensor which can be put on the right-hand side of the classical Einstein equations. [[35], p.2748]

As we will describe, that turns out to have an been unduly pessimistic assessment, albeit understandable given the problematic nature of the particle concept. However, the full solution to the problem really requires us to break away from the conventional Fock space picture of QFT in favour of an algebraic approach.

5 Algebraic approach to quantization

The algebraic approach to QFT in CST resolves the problems encountered in section 4.7 by separating out the invariant elements of the theory—the algebraic relations obeyed by the smeared fields—from the problem of finding concrete Hilbert space representations. A number of interesting questions can be addressed without ever passing to a Hilbert space–see, for example, section 6.4. For a survey of the extensive subject of the algebraic QFT in *flat* spacetime, see Haag's monograph [36].

5.1 The algebra

In section 3.4 we identified the basic relations that should be obeyed by smeared Klein–Gordon fields:

- $f \mapsto \Phi(f)$ is complex-linear;
- $\Phi(f)^* = \Phi(\overline{f})$ for all $f \in C_0^\infty(M)$
- $\Phi(Pf) = 0$ for all $f \in C_0^\infty(M;\mathbb{R})$ for all $f \in C_0^\infty(M)$
- $[\Phi(f), \Phi(f')] = iE(f, f')\mathbb{1}$ for all $f, f' \in C_0^\infty(M)$.

In addition, we expect to be able to form linear combinations of finite products of smeared fields and their adjoints. Mathematically, this means that we want a $*$-algebra generated by some objects denoted $\Phi(f)$ (labelled by $f \in C_0^\infty(M)$) and a unit $\mathbb{1}$, satisfying the above relations. The symbol $\Phi(f)$ should not be thought of as the integral of some underlying $\Phi(x)$ against $f(x)$. In essence it is just f in disguise.

There is a standard construction of an algebra $\mathcal{A}(\boldsymbol{M})$ with the required properties. (We adopt the convention that $\boldsymbol{M}$ denotes the manifold, metric and time-orientation [and any other objects we require, e.g., orientation, spin-structure...].) This algebra consists of finite linear combinations of finite products of the $\Phi(f)$, $\Phi(f)^*$ and $\mathbb{1}$, counting two such elements as equal if they can be manipulated into a common form using the relations above. Note that there is no Hilbert space, nor have we invoked any coordinates or symmetries.

In more detail, the construction proceeds as follows:

- Start with the set of generators $\mathcal{F} = \{\Phi(f) : f \in C_0^\infty(M)\}$.

- Generate a free unital $*$-algebra $\mathcal{A}$ over $\mathcal{F}$, which means that $\mathcal{A}$ consists of finite linear combinations of finite products of the $\Phi(f)$, the $\Phi(f)^*$ and $\mathbb{1}$.[12]

- Let $\mathscr{I}$ be the subset of $\mathcal{A}$ consisting of all finite linear combinations of elements of the form ABC where $A, C \in \mathcal{A}$ and B or its adjoint is one of

$$\begin{aligned}
&\Phi(\lambda f + \mu f') - (\lambda\Phi(f) + \mu\Phi(f')) \\
&\Phi(f)^* - \Phi(\overline{f}) \\
&\Phi(Pf) \\
&\Phi(f)\Phi(f') - \Phi(f')\Phi(f) - iE(f, f')\mathbb{1}
\end{aligned}$$

  for $f, f' \in C_0^\infty(M)$, $\lambda, \mu \in \mathbb{C}$. By construction, $\mathscr{I}$ is a linear subspace of $\mathcal{A}$ that is invariant under the adjoint and also products by elements of $\mathcal{A}$ from left or right. That is $\mathscr{I}$ is a **$*$-ideal** in $\mathcal{A}$.

- The quotient vector space $\mathcal{A}/\mathscr{I}$, consisting of equivalence classes in $\mathcal{A}$ so that

$$[A] = [B] \qquad \iff \qquad A - B \in \mathscr{I},$$

  inherits algebraic operations from $\mathcal{A}$, e.g., we define

$$[A][B] = [AB]$$

  and check that it is well-defined: if $[A] = [A']$, $[B] = [B']$ then

$$A'B' = (A + I)(B + J) = AB + \underbrace{IB + AJ + IJ}_{\in \mathscr{I}}$$

  so $[A'B'] = [AB]$.

- The algebra $\mathcal{A}(\boldsymbol{M})$ is then defined to be the quotient $\mathcal{A}/\mathscr{I}$, equipped with the operations just mentioned. We write $\Phi(f)$ for the equivalence class more properly written $[\Phi(f)]$.

- Summarising, every element in $\mathcal{A}(\boldsymbol{M})$ is a finite linear combinations of finite products of the $\Phi(f)$, the $\Phi(f)^*$ and $\mathbb{1}$; moreover, the quotient construction ensures that all the desired relations are obeyed.

Of course, it is also necessary to know that the resulting algebra is nontrivial. One approach to this is to appeal to the existence of nontrivial representations. A direct method is also possible: one may use the defining relations to show that $\mathcal{A}(\boldsymbol{M})$ is isomorphic *as a vector space* to the symmetric algebra[13] over the solution space $\mathsf{Sol}(\boldsymbol{M})$ on $\boldsymbol{M}$.

[12] This seems circular: how can we talk about products before we have the algebra? The point is that $\mathcal{F} \cup \mathcal{F}^* \cup \{\mathbb{1}\}$ should be thought of as an alphabet, and any finite word [string of letters] is what we mean by a finite product; concatenation of words provides a product. Linear combinations are dealt with in a similar way.

[13] "Symmetric algebra" = "free unital commutative complex algebra".

Finally, we mention another commonly-employed algebraic formulation of the real scalar field based on the so-called **Weyl algebra** Weyl($\boldsymbol{M}$). This is defined as follows: let $\mathcal{W}$ be the free unital $*$-algebra generated by elements $\{W(\phi) : \phi \in \mathsf{Sol}(\boldsymbol{M})\}$ and form a quotient (in the same way as above) of $\mathcal{W}$ by the relations

$$W(-\phi) = W(\phi)^*$$
$$W(\phi)W(\phi') = e^{-i\sigma(\phi,\phi')/2}W(\phi+\phi')$$

for all $\phi, \phi' \in \mathsf{Sol}(\boldsymbol{M})$. Here, $\sigma : \mathsf{Sol}(\boldsymbol{M}) \times \mathsf{Sol}(\boldsymbol{M}) \to \mathbb{C}$ is defined so that $\sigma(Ef, Ef') = E(f, f')$ for all $f, f' \in C_0^\infty(\boldsymbol{M}))$ (cf. Eq. (4)): it is the complexification of the symplectic form on the space of real Klein–Gordon solutions. The Weyl algebra Weyl($\boldsymbol{M}$) is the unique completion of this quotient as a C^*-algebra. The distinction between Weyl($\boldsymbol{M}$) and $\mathcal{A}(\boldsymbol{M})$ is primarily technical: any 'sufficiently nice' representation π : Weyl $\to B(\mathscr{H})$ in a Hilbert space $\mathscr{H}$ induces a representation $\widetilde{\pi}$ of $\mathcal{A}(\boldsymbol{M})$ as unbounded operators on $\mathscr{H}$, so that

$$\pi(W(Ef)) = e^{i\widetilde{\pi}(\Phi(f))}$$

for real-valued test functions f. The main advantages of the Weyl algebra are: (a) that it is generally easier to work with bounded rather than unbounded operators; (b) expectation values of the $W(\phi)$ take a simple form in quasifree states (cf. Eq. (12)).

5.2 States and the GNS representation

Self-adjoint elements $[A^* = A]$ of $\mathcal{A}(\boldsymbol{M})$ should play the role of observables. However, this is rather empty without a rule for turning observables into expectation values, in other words, notion of a state.

Definition 5.1 *A state on $\mathcal{A}(\boldsymbol{M})$ is a linear map $\omega : \mathcal{A}(\boldsymbol{M}) \to \mathbb{C}$ obeying*

$$\omega(\mathbb{1}) = 1 \qquad \textit{normalisation}$$
$$\forall A \in \mathcal{A}(\boldsymbol{M}),\ \omega(A^*A) \geq 0 \qquad \textit{positivity.}$$

Expectation values

$$\omega_n(f_1, \ldots, f_n) \stackrel{\text{def}}{=} \omega(\Phi(f_1)\Phi(f_2)\cdots\Phi(f_n))$$

are called n-point functions. It is clearly sufficient to specify the n-point functions to fix ω. Perhaps reassuringly, given a state ω we may regain a Hilbert space setting using the so-called GNS construction (Gel'fand, Naimark, Segal):

- Define
  $$\mathscr{I}_\omega = \{A \in \mathcal{A}(\boldsymbol{M}) : \omega(A^*A) = 0\}.$$
  We claim that $\mathscr{I}_\omega$ is a vector subspace and left-ideal of $\mathcal{A}(\boldsymbol{M})$. To see this we use positivity of ω to deduce the Cauchy–Schwarz inequality
  $$|\omega(A^*B)|^2 \leq \omega(A^*A)\omega(B^*B).$$

Then

$$\omega((BA)^*(BA))^2 = \omega(A^*B^*BA)^2 \leq \omega(A^*A)\omega((B^*BA)^*(B^*BA)),$$

so $A \in \mathscr{I}_\omega$ implies $BA \in \mathscr{I}_\omega$ for any B. Similarly we may show that $\mathscr{I}_\omega$ is a vector space by expanding $\omega((\lambda A + \mu B)^*(\lambda A + \mu B))$ and using Cauchy–Schwarz.

- Defining the vector space quotient $\mathscr{D}_\omega = \mathcal{A}(\boldsymbol{M})/\mathscr{I}_\omega$, we claim that

  $$\langle [A] \mid [B] \rangle = \omega(A^*B),$$

  defines an inner product on $\mathscr{D}_\omega$. (One must check that this is well defined [NB $\mathscr{I}_\omega^*$ is a *right* ideal] and obeys the axioms of an inner product.)

- Define $\mathscr{H}_\omega$ to be the Hilbert space completion of $\mathscr{D}_\omega$ in the inner product just constructed, and write $\Omega_\omega = [\mathbb{1}]$.

- For each $A \in \mathcal{A}(\boldsymbol{M})$ define an operator $\pi_\omega(A) : \mathscr{D}_\omega \to \mathscr{D}_\omega$ by

  $$\pi_\omega(A)[B] = [AB],$$

  which is well-defined by the left-ideal property of $\mathscr{I}_\omega$. It is straightforward to verify that

  $$\begin{aligned} \pi_\omega(\mathbb{1}) &= \mathbb{1}_{\mathscr{H}_\omega}|_{\mathscr{D}_\omega} \\ \pi_\omega(AB) &= \pi_\omega(A)\pi_\omega(B) \\ \pi_\omega(\lambda A + \mu B) &= \lambda\pi_\omega(A) + \mu\pi_\omega(B) \\ \pi_\omega(A)^*|_{\mathscr{D}_\omega} &= \pi_\omega(A^*). \end{aligned}$$

  Thus π_ω is a $*$-representation of $\mathcal{A}(\boldsymbol{M})$ among the unbounded operators defined on $\mathscr{D}_\omega$ in $\mathscr{H}_\omega$. Moreover,

  $$\omega(A) = \langle \Omega_\omega \mid \pi_\omega(A)\Omega_\omega \rangle$$

  so Ω_ω represents the state in the representation.

The quadruple $(\mathscr{H}_\omega, \mathscr{D}_\omega, \Omega_\omega, \pi_\omega)$ is the **GNS representation** of $\mathcal{A}(\boldsymbol{M})$ induced by the state ω. It is unique up to unitary equivalence.

Example: The Fock spaces constructed in earlier sections are easily seen to carry a $*$-representation of the algebra $\mathcal{A}(\boldsymbol{M})$ for the spacetime(s) concerned. The expectation values in the Fock vacuum state Ω induce a state ω_Ω on $\mathcal{A}(\boldsymbol{M})$ whose corresponding GNS representation $(\mathscr{H}_{\omega_\Omega}, \mathscr{D}_{\omega_\Omega}, \Omega_{\omega_\Omega}, \pi_{\omega_\Omega})$ is (unitarily equivalent to) the Fock representation so that the GNS vector Ω_{ω_Ω} is equal (under the equivalence) to Ω.

To summarise this section: the GNS representation permits us to return to a Hilbert space representation once we have chosen a state. Of course, the representations induced by different states may turn out not to be unitarily equivalent; the advantage of the algebraic approach is that we also have an arena [namely, the algebra] in which such states can be treated in a democratic fashion.

6 Microlocal analysis and the Hadamard condition

The set of all states on the algebra of the Klein–Gordon field is too large for many purposes. For example, there are states whose n-point functions are not continuous in the test functions, or are insufficiently regular to permit the construction of Wick or time-ordered products, e.g., to calculate the stress-energy tensor or for perturbation theory. This section discusses the class of Hadamard states that has come to be accepted as the 'right' class of states for the Klein–Gordon field. In particular, we explain how it may be formulated as a microlocal spectrum condition using techniques drawn from microlocal analysis. Rather than follow the historical development, we proceed in reverse, indicating how the microlocal spectrum condition emerges as a natural necessary condition on states for the construction of Wick powers, and then describing results of Radzikowski that show that it is also sufficient for this purpose, and picks out the Hadamard class.

6.1 Motivation: Wick powers

In the conventional Minkowski space QFT, the Wick square may be defined by **point-splitting** as follows. Beginning with

$$\Phi(x)\Phi(x') = \int \frac{d^3\underline{k}}{(2\pi)^3}\frac{d^3\underline{k}'}{(2\pi)^3}\frac{1}{2\sqrt{\omega\omega'}}\left(a(\underline{k})\mathrm{e}^{-ikx} + a(\underline{k})^*\mathrm{e}^{ikx}\right)\left(a(\underline{k}')\mathrm{e}^{-ik'x'} + a(\underline{k}')^*\mathrm{e}^{ik'x'}\right),$$

we apply normal ordering, replacing $a(\underline{k})a(\underline{k}')^*$ by $a(\underline{k}')^*a(\underline{k})$. This has the same effect as subtracting

$$[a(\underline{k}), a(\underline{k}')] = \delta^3(\underline{k}-\underline{k}')\mathbb{1},$$

so

$$:\!\Phi(x)\Phi(x')\!: = \Phi(x)\Phi(x') - \underbrace{\int \frac{d^3\underline{k}}{(2\pi)^3}\frac{1}{2\omega}\mathrm{e}^{-ik(x-x')}}_{\omega_2^{\mathrm{vac}}(x,x')}\mathbb{1}$$

and hence

$$\omega(:\!\Phi(x)\Phi(x')\!:) = \omega_2(x,x') - \omega_2^{\mathrm{vac}}(x,x').$$

The expectation value $\omega(:\!\Phi^2\!:(x))$ is defined by taking the limit $x' \to x$, a precondition for which is that the right-hand side is continuous. To ensure the existence of Wick squares involving arbitrary derivatives of Φ, it is reasonable to demand that that the right-hand side should be smooth for any 'physically acceptable' state ω. It follows that the any two physically acceptable states have two-point functions whose difference is smooth. This motivates the study of the singularity structure of ω_2^{vac}, which we undertake using ideas taken from microlocal analysis.

6.2 The wavefront set

Fourier analysis provides a fundamental duality between smoothness and decay: smooth functions have rapidly decaying Fourier transforms, and vice versa. The fundamental idea underlying microlocal analysis is that decay properties of the Fourier transform of a distribution can be used to obtain detailed information about its singular structure. A general reference for this section is [44], particularly chapter 8.

We begin with some examples, recalling our (nonstandard) convention used for Fourier transform, namely

$$\widehat{f}(k) = \int d^n x \, e^{ik\cdot x} f(x).$$

Examples

a) If $f \in C_0^\infty(\mathbb{R}^n)$ then

$$\left(1 + |k|^{2m}\right) \left|\widehat{f}(k)\right| = |(1 + (-\triangle)^m f)^\wedge(k)| \leq \int d^n x \, |(1 + (-\triangle)^m f)(x)| < \infty.$$

So for each N, there exists a constant C_N such that

$$\left|\widehat{f}(k)\right| \leq \frac{C_N}{1 + |k|^N} \qquad \text{as } k \to \infty$$

(this is what we mean by 'rapid decay'.)

b) The δ-distribution at the origin has Fourier transform $\widehat{\delta}(k) = 1$, which exhibits no decay at ∞.

c) The distribution $T \in \mathscr{D}'(\mathbb{R})$ defined by

$$T(f) = \lim_{\varepsilon \to 0^+} \int \frac{f(s)}{s - i\varepsilon} ds$$

has Fourier transform

$$\widehat{T}(k) = \lim_{\varepsilon \to 0^+} \int \frac{\mathrm{e}^{iks}}{s - i\varepsilon} ds = 2\pi i \Theta(k),$$

which decays as $k \to -\infty$ but not as $k \to +\infty$.

The wavefront set localises information of this type both in x-space and on the "sphere at ∞" on k-space.

Definition 6.1 *A) If $u \in \mathscr{D}'(\mathbb{R}^n)$, a pair $(x, k) \in \mathbb{R}^n \times (\mathbb{R}^n \backslash \{0\})$ is a* regular direction *for u if there exist*

i) $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\phi(x) \neq 0$

ii) a conic neighbourhood V of k

iii) constants C_N, $N \in \mathbb{N}$

so that

$$\left|\widehat{\phi u}(k)\right| < \frac{C_N}{1 + |k|^N} \qquad \forall k \in V, N \in \mathbb{N}$$

i.e., $\widehat{\phi u}$ decays rapidly as $k \to \infty$ in V.

B) The wavefront set *of u is defined to be*

$$\mathit{WF}(u) = \{(x,k) \in \mathbb{R}^n \times (\mathbb{R}^n \backslash \{0\}) : (x,k) \textit{ is } \text{not } \textit{a regular direction for } u\}.$$

Examples

a) If $f \in C^\infty(\mathbb{R}^n)$, then $\mathrm{WF}\,(f) = \emptyset$.

b) $\mathrm{WF}\,(\delta) = \{(0,k) \in \mathbb{R}^2 : k \neq 0\}$. (Note that $\widehat{\phi\delta}(k) = \phi(0)$, so (x,k) is a regular direction for $x \neq 0$ as we may then choose ϕ with $\phi(x) \neq 0$, $\phi(0) = 0$).

c) $\mathrm{WF}\,(T) = \{(0,k) \in \mathbb{R}^2 : k > 0\}$ (exercise!).

The wavefront set has many natural and useful properties. For our purposes, the most important are the following:

- $\mathrm{WF}(u) = \phi \Longleftrightarrow u \in C^\infty(\mathbb{R}^n)$.
- $\mathrm{WF}(\lambda u + \mu v) \subset \mathrm{WF}(u) \cup \mathrm{WF}(v)$ for $u, v \in \mathscr{D}'(\mathbb{R}^n)$, $\lambda, \mu \in \mathbb{C}$.
- If P is any partial differential operator with smooth coefficients, then

  $$\mathrm{WF}(Pu) \subset \mathrm{WF}(u) \subset \mathrm{WF}(Pu) \cup \mathrm{Char}\, P,$$

  for any $u \in \mathscr{D}'(\mathbb{R}^n)$, where $\mathrm{Char}\, P$ is the **characteristic set** of P. To define the characteristic set, let m be the order of P, i.e., the least $m \in \mathbb{N}_0$ so that P may be written in the form $P = \sum_{|\alpha| \leq m} a_\alpha(x)(iD)^\alpha$ where α is a multiindex. The **principal symbol** of P is the smooth function on $\mathbb{R}^n \times \mathbb{R}^n$ given by

  $$p_m(x,k) = \sum_{|\alpha|=m} a_\alpha(x) k^\alpha$$

  and the characteristic set is

  $$\mathrm{Char}\, P = \{(x,k) \in \mathbb{R}^n \times (\mathbb{R}^{n*} \setminus \{0\}) : p_m(x,k) = 0\}.$$

- **Propagation of Singularities**: $\mathrm{WF}(u) \setminus \mathrm{WF}(Pu)$ is invariant under the Hamiltonian flow generated by p_m.
- Under coordinate changes, WF and Char transform as subsets of the cotangent bundle: given a diffeomorphism φ, define $\varphi^* u$ by $(\varphi^* u)(f) = u(f \circ \varphi)$. Then

  $$\mathrm{WF}(u) = \{(x, \xi D\varphi|_x) : (\varphi(x), \xi) \in \mathrm{WF}(\varphi^* u)\};$$

  similarly, setting $(P_\varphi f) \circ \varphi = P(f \circ \varphi)$, we have

  $$(p_\varphi)_m(\varphi(x), \xi) = p_m(x, \xi D\varphi|_x).$$

  In particular, we may extend the wavefront set and characteristic set to distributions and partial differential operators defined on manifolds; both are subsets of the cotangent bundle.

Examples:

1. Let P be the Klein–Gordon operator $P = \Box_g + m^2 + \xi R$ on a spacetime (M, g). The principal symbol is easily seen to be

$$p_2(x, \xi) = -g^{ab}(x)\xi_a\xi_b$$

and so the characteristic set is

$$\mathrm{Char}\, P = \mathcal{N},$$

where $\mathcal{N}$ is the bundle of nonzero null covectors on M:

$$\mathcal{N} = \{(x, \xi) \in T^*M : \xi \text{ a non-zero null at p}\}.$$

Hence the wavefront set of any (distributional) solution to $Pu = 0$ obeys

$$\mathrm{WF}(u) \subset \mathcal{N};$$

moreover, $\mathrm{WF}(u)$ is invariant under the Hamiltonian evolution $\lambda \mapsto (x(\lambda), \xi(\lambda)) \in T^*M$ given by the 'Hamiltonian' $p_2(x, \xi)$, namely

$$\begin{aligned}
\dot{x}^a &= -g^{ab}\xi_b \\
\dot{\xi}_c &= -(\nabla_c g^{ab})\xi_a\xi_b = 0,
\end{aligned}$$

which simply asserts that $\dot{x}^a$ is parallel transported along the curve $\lambda \mapsto x(\lambda)$ and that ξ_a is cotangent to this curve. Thus if $(x, \xi) \in \mathrm{WF}\,(u)$, the wavefront set contains every point $(x(\lambda), \xi(\lambda))$ for $\lambda \in \mathbb{R}$, where $x(\lambda)$ is the null geodesic through x with tangent $\xi^\sharp$ and $\xi(\lambda)$ is the parallel transport of ξ along $x(\lambda)$.

2. Now consider Klein–Gordon bisolutions, i.e., $F \in \mathscr{D}'(M \times M)$ such that

$$(P \otimes 1)F = (1 \otimes P)F = 0.$$

Now the operator $P \otimes 1$ has principal symbol

$$p(x, \xi; x', \xi') = g^{ab}(x)\xi_a\xi_b$$

and characteristic set

$$\mathrm{Char}\, P \otimes 1 = \mathcal{N}_0 \times \dot{T}^*M$$

where $\mathcal{N}_0$ is the bundle of (possibly zero) null covectors on M (i.e., $\mathcal{N}$, with the zero covector added at each point) and $\dot{T}^*M$ is the cotangent bundle of M with the zero section[14] deleted. Similarly, $1 \otimes P$ has principal symbol

$$p'(x, \xi; x', \xi') = g^{ab}(x')\xi'_a\xi'_b$$

and characteristic set

$$\mathrm{Char}\, 1 \otimes P = \dot{T}^*M \times \mathcal{N}_0.$$

The bisolution F therefore has wavefront set with upper bound

$$\mathrm{WF}\,(F) \subset \left(\mathcal{N}_0 \times \dot{T}^*M\right) \cap \left(\dot{T}^*M \times \mathcal{N}_0\right) \subset \mathcal{N}_0 \times \mathcal{N}_0.$$

[14]The zero section consists of all elements $(x, 0) \in T^*M$.

6.3 Back to QFT

We now return to our motivating application: the definition of Wick powers and their derivatives. Beginning with Minkowski space (M, η), we have already seen (in Section 6.1) that a *necessary* condition on a state ω for it to assign well-defined expectation values to all such objects was that

$$\omega_2 - \omega_2^{\rm vac} \in C^\infty(M \times M),$$

where $\omega_2^{\rm vac}$ is the Minkowski vacuum state. It follows from the properties of the wavefront set stated in section 6.2 that

$$\mathrm{WF}\,(\omega_2 - \omega_2^{\rm vac}) = \emptyset$$

and hence

$$\mathrm{WF}\,(\omega_2) = \mathrm{WF}(\omega_2^{\rm vac}).$$

Moreover, we know that

$$\mathrm{WF}\,(\omega_2^{\rm vac}) \subset \mathcal{N}_0 \times \mathcal{N}_0 \tag{14}$$

because all two-point functions are bisolutions. Now for $\phi(x_1, x_2) = \phi_1(x_1)\phi_2(x_2)$ and $\phi_i \in C_0^\infty(M)$, we may compute

$$\widehat{\phi\omega_2^{\rm vac}}(l, l') = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega}\widehat{\phi}_1(l-k)\widehat{\phi}_2(l'+k)$$

with future pointing, on-shell k. As the functions ϕ_i are smooth, their Fourier transforms decay rapidly as their arguments become large. The main contribution to the integral therefore arises from regions of k where $l-k$ and $l'+k$ are simultaneously small, i.e., ℓ must be near to the future pointing on-shell covector k, and ℓ' must be near $-k$. Arguing in this way, it is not hard to see that the integral will tend to zero as $(l, l') \to \infty$ in open conic neighbourhoods of $\mathbb{R}^4_- \times \mathbb{R}^4$ and in $\mathbb{R}^4 \times \mathbb{R}^4_+$, where $\mathbb{R}^4_\pm$ is the half-space in which $\pm k_0 \geq 0$. Thus $(x_1, k_1; x_2, k_2)$ is a regular direction if either (i) k_1 is zero or past-directed; or, (ii) k_2 is zero or future-directed. Putting this together with the upper bound (14), we conclude that

$$\mathrm{WF}\,(\omega_2^{\rm vac}) \subset \mathcal{N}^+ \times \mathcal{N}^-, \tag{15}$$

where

$$\mathcal{N}^\pm = \{(p, \xi) \in \mathcal{N} : \xi \text{ is future}(+)/\text{past}(-)\text{ directed}\}.$$

In a general curved spacetime, it must still be the case that $\mathrm{WF}(\omega_2) \subset \mathcal{N}_0 \times \mathcal{N}_0$. The minimal extension of our necessary condition on ω for the existence of Wick powers is to elevate (15) to a general requirement.

Definition 6.2 *A state ω obeys the* **Microlocal Spectrum Condition** *(μSC)*[15] *if*

$$WF(\omega_2) \subset \mathcal{N}^+ \times \mathcal{N}^-.$$

[15] The term microlocal spectrum condition was introduced in [7] with an apparently stronger definition; see the remarks at the end of this section.

In particular, this asserts that the 'singular behaviour' of the two-point function is positive-frequency in the first slot and negative-frequency in the second. We have already argued that the Minkowski vacuum obeys the μSC; it is also true that ground and thermal states on various classes of stationary spacetime, including those constructed in section 4 also satisfy the μSC [45, 61, 66]. (In relation to thermal states, the key point is that negative frequency contributions to the first slot of the thermal two-point functions are exponentially suppressed.) Furthermore, our discussion above makes it very plausible that μSC should be a necessary condition for the existence of expectation values of Wick powers. At first sight, however, it seems a long way from being sufficient. Remember that we need a condition on states that guarantees that differences between 2-point functions are smooth. But even if two distributions have the same wavefront set, their difference is not necessarily smooth ($\mathrm{WF}\,(\delta) = \mathrm{WF}\,(2\delta)$, for instance). With this in mind, it is remarkable that the μSC actually does what we need. The following result is due to Radzikowski [59].

Theorem 6.3 *If ω and ω' obey the μSC then*

$$\omega_2 - \omega_2' \in C^\infty(M \times M)$$

i.e., the μSC determines an equivalence of class of states under equality of two-point functions modulo C^∞.

As mentioned, it is surprising that such a result can be true. The key point is that, while the antisymmetric parts of ω_2 and ω_2' are both equal to $\frac{1}{2}iE$, $\mathrm{WF}(\omega_2)$ is not the whole of $\mathrm{WF}(E)$, which intersects both $\mathcal{N}^+ \times \mathcal{N}^-$ and $\mathcal{N}^- \times \mathcal{N}^+$. Accordingly, the singularities in the symmetric part must precisely cancel the unwanted singular directions in $\mathrm{WF}\,(E)$, which is how the microlocal spectrum condition does, after all, fix the singular structure of the two-point function.

It follows from Theorem 6.3 that all two-point functions of states obeying μSC must have equal wavefront sets. The universal nature of the antisymmetric part of the two-point function is universal also allows us to fix the wavefront set of the two-point function in the following way.

Lemma 6.4 *If ω obeys the μSC then $WF(\omega_2) = WF(E) \cap (\mathcal{N}^+ \times \mathcal{N}^-)$.*

Proof: Define $\tilde{\omega}_2(x, x') = \omega_2(x', x)$, so $\mathrm{WF}\,(\tilde{\omega}_2) \subset \mathcal{N}^- \times \mathcal{N}^+$ by the μSC. But

$$\begin{aligned} \mathrm{WF}\,(E) &\subset \mathrm{WF}\,(\omega_2) \cup \mathrm{WF}\,(\tilde{\omega}_2) && [\text{as } iE = \omega_2 - \tilde{\omega}_2] \\ &\subset \mathrm{WF}\,(\tilde{\omega}_2) \cup \mathrm{WF}\,(E) && [\text{as } \omega_2 = \tilde{\omega}_2 + iE] \end{aligned}$$

so, using again the fact that $\mathrm{WF}\,(\tilde{\omega}_2) \subset \mathcal{N}^- \times \mathcal{N}^+$

$$\mathrm{WF}\,(E) \subset \mathrm{WF}(\omega_2) \cup (\mathcal{N}^- \times \mathcal{N}^+) \subset (\mathcal{N}^- \times \mathcal{N}^+) \cup \mathrm{WF}\,(E),$$

and we take intersections with $\mathcal{N}^+ \times \mathcal{N}^-$ to obtain the required result. $\square$

The wavefront set of E is known from work of Duistermaat and Hörmander on distinguished parametrices. This permits us to give a final form of the wavefront set of a Hadamard 2-point function:

$$\mathrm{WF}\,(\omega_2) = \{(p,\xi;p,-\xi') \in \dot{T}^*(M\times M):\ (p,\xi)\sim(p',\xi') \text{ and } \xi\in\mathcal{N}^+\}, \tag{16}$$

where $\dot{T}^*(M\times M)$ is the cotangent bundle of $M\times M$ with its zero section[16] deleted and the equivalence relation $\sim$ is defined so that $(p,\xi)\sim(p',\xi')$ if and only if either

- there is a null geodesic γ connecting p and p', so that ξ is parallel to $\dot\gamma^\flat$ at p, and ξ' is the parallel transport of ξ to p' (and necessarily parallel to $\dot\gamma^\flat$ at p'); or,
- $p=p'$ and $\xi=\xi'$.

Eq. (16) is the form that Radzikowski stated as his 'wavefront set spectral condition'. In fact, Radzikowski proved much more than Theorem 6.3 in [59].

Theorem 6.5 (Radzikowski) *The μSC is equivalent to the* **Hadamard condition***.*

The Hadamard condition became prominent during the mid-to-late 1970's as the basis for the point-splitting technique for calculating expected values of the stress-energy tensor. In fact the early attempts were not entirely successful, as they did not give the trace anomaly produced by other regularization methods, owing to an error identified and fixed by Wald [69]. Moreover, the Hadamard condition was a bit unwieldy and, indeed, a fully precise version was not given until the 1991 paper of Kay and Wald [49].

We will not give full details here, but restrict to an outline, referring to [49, 71, 42] for the details. The essential idea is that in any convex normal and causally convex neighbourhood $\mathcal{O}$, a **local Hadamard parametrix** $H_{\mathcal{O}}\in\mathscr{D}'(\mathcal{O}\times\mathcal{O})$ may be constructed using the local geometry and the Klein–Gordon operator. This takes the basic form (in four dimensions)

$$H_{\mathcal{O}}(x,x') = \frac{U(x,x')}{4\pi^2\sigma_+(x,x')} + V(x,x')\log(\sigma_+/\ell^2)$$

where U and V are smooth, $f(\sigma_+)$ is a regularization of $f(\sigma)$, with σ the signed square geodesic separation of x and x'. The parameter ℓ is a length scale, necessary for dimensional reasons. The functions U and V are defined using ℓ, the local geometry of $\mathcal{O}$ and the Klein–Gordon operator, along with the condition that $U(x,x)=1$. Note that $H_{\mathcal{O}}$ has the same form as the Minkowski two-point function discussed in section 4.4.

The $H_{\mathcal{O}}$ are often presented in the **Hadamard series** form, in which U and V are expanded as series in σ with coefficients determined by Hadamard recursion relations. This series does not converge in general, a problem which we resolve (as in section 5.2 in [42]) by inserting cut-off functions into the series to enforce convergence without disturbing the singularity structure. An alternative approach is to truncate the series,

[16] I.e., elements of the form $(p,0;p',0)$.

which is the method adopted in [49] and much of the literature. For later purposes it is also useful to assume that the antisymmetric part of $H_{\mathcal{O}}$ is fixed so that

$$H_{\mathcal{O}}(x, x') - H_{\mathcal{O}}(x', x) = iE(x, x')$$

just as is the case for a two-point function.

The essence of the Hadamard condition may be formulated as follows (see [49] for the precise version): A state ω on $\mathcal{A}(\boldsymbol{M})$ is Hadamard if $\omega_2 - H_{\mathcal{O}}$ is smooth for a set of $\mathcal{O}$ that provide a suitable cover of a Cauchy surface in $\boldsymbol{M}$. It then follows that $\omega_2 - H_{\mathcal{O}}$ is smooth for every $\mathcal{O}$.

Although the bidistribution $H_{\mathcal{O}}$ does depend on the way one chooses the cutoffs mentioned above, the diagonal values of $\omega_2 - H_{\mathcal{O}}$ and its derivatives are independent of these choices. We can now give a straightforward prescription for defining the Wick square at any $x \in \mathcal{O}$ by point-splitting:

$$\omega(\boldsymbol{:}\Phi^2\boldsymbol{:}(x)) = (\omega_2 - H_{\mathcal{O}})(x, x).$$

The stress-energy tensor is a little more subtle and we refer to [71] for a detailed discussion and original literature, restricting ourselves to a sketch. Let $\mathcal{T}$ be a differential operator that maps smooth functions on $U \times U$ to smooth bi-covector fields on $U \times U$, with the property that $\mathcal{T}(\phi \otimes \phi)(x, x)$ is the classical stress-energy tensor of any Klein–Gordon solution ϕ. Applying $\mathcal{T}$ to $\omega_2 - H_{\mathcal{O}}$ and bringing the points together, we obtain a rank-2 covariant tensor field $x \mapsto (\mathcal{T}(\omega_2 - H_{\mathcal{O}}))(x, x)$ on U. It turns out that although this tensor field is not necessarily conserved, the problem can be fixed by subtracting a local geometrical term of the form Qg_{ab}, and can be avoided altogether by a clever choice of $\mathcal{T}$ [54].

Remarks:

1. As mentioned above, the final stress-energy tensor exhibits a **trace anomaly**: even if the classical theory is conformally invariant (e.g., $m = 0$ and $\xi = 1/6$ in four dimensions) the trace of the renormalised stress-energy tensor is nonvanishing, in contrast to the classical situation.

2. The length scale ℓ provides a residual freedom in the definition of $H_{\mathcal{O}}$ and hence a one-parameter family of renormalized Wick squares and stress-energy tensors. In the case of the stress-energy tensor there is a wider four-parameter family of alternatives, differing by conserved symmetric local curvature tensors that are at most quadratic in the curvature, which obey axioms for renormalization due to Wald and arise in other renormalization schemes. (See [71]; the original references [68, 69] include an additional axiom later dropped from the list.) For similar issues in the case of higher Wick powers see [42], and [10] for a proposal to fix these freedoms using thermodynamic data). These freedoms are often called ambiguities, but in my view are better regarded as relating to different extensions of the same underlying free theory, distinguished by the renormalized products of the underlying fields.

3. As the following quotation, taken from the 1978 paper of Fulling, Sweeny and Wald [32], makes clear, the introduction of the Hadamard condition was a spur to the development of the algebraic approach to QFT in CST:

   > All these considerations suggest that the validity of [the Hadamard condition] be regarded as a basic criterion for a "physically reasonable" state, perhaps even as the definition of that phrase. This raises the possibility of constructing quantum states from two-point distribution solutions of the field equation by a procedure of the Wightman or GNS type... ...bypassing the quantization of normal modes in a Fock space.

4. We have only discussed regularity of the 2-point function. In some references, the term microlocal spectrum condition is defined as a condition on all n-point functions of the form
$$\mathrm{WF}\,(\omega_n) \subset \Gamma_n$$
where the Γ_n are particular subsets of $T^*M^{\times n}$. This condition was introduced in [7], where it is also shown to be satisfied by all quasifree Hadamard states. Very recently, Sanders has proved that this apparently more general condition is actually equivalent to the μSC in the form we have stated; and, moreover, that all states obeying μSC have smooth truncated n-point functions for $n \neq 2$ [63]. One may interpret this as saying that all Hadamard states are 'microlocally quasifree'; it also shows that the class of (not necessarily quasifree) Hadamard states is precisely the 'state space of perturbative QFT' studied by Hollands & Ruan in [41], and previously identified as a plausible class of interest by Kay [47].

5. Finally, we mention a variation on the theme. For some purposes it is sufficient only to require the two-point functions to agree with $H_{\mathcal{O}}$ modulo some Sobolev space, rather than modulo C^∞. This leads to the microlocal study of **adiabatic states** [46].

6.4 Quantum (energy) inequalities

As an application of the ideas presented in the last two chapters we give a brief discussion of quantum inequalities, aiming to show how general results can be obtained using the algebraic properties of the quantum field and the microlocal spectrum condition, without ever using a Hilbert space representation. We start from the fact that, although the square of a classical real scalar field is, everywhere nonnegative, the same is not true of the Wick square we have just constructed.

In four-dimensional Minkowski space, for example, we have
$$:\!\Phi^2\!:(f)\Omega = \int \frac{d^3\underline{k}}{(2\pi)^3}\frac{d^3\underline{k}'}{(2\pi)^3}\frac{1}{2\sqrt{\omega\omega'}}\widehat{f}(k+k')a(\underline{k})^*a(\underline{k}')^*\Omega$$
for any compactly supported test function f; it is obvious that $\langle \Omega \mid :\!\Phi^2\!:(f)\Omega\rangle = 0$, and a short calculation gives
$$\|:\!\Phi^2\!:(f)\Omega\|^2 = \int \frac{d^3\underline{k}}{(2\pi)^3}\frac{d^3\underline{k}'}{(2\pi)^3}\frac{|\widehat{f}(k+k')|^2}{2\omega\omega'},$$

which is nonzero unless f is identically zero[17]. The observable $:\Phi^2:(f)$ therefore has vanishing expectation value in the state Ω, but does not annihilate Ω. Standard variational arguments imply that $:\Phi^2:(f)$ must have some negative spectrum–indeed, if we write

$$\psi_\alpha = \cos\alpha\Omega + \sin\alpha :\Phi^2:(f)\Omega$$

(assuming f is chosen so $\| :\Phi^2:(f)\Omega\| = 1$) it is easy to calculate

$$\langle \psi_\alpha \mid :\Phi^2:(f)\psi_\alpha\rangle = 2\alpha + O(\alpha^2)$$

giving negative expectation values for sufficiently small $\alpha < 0$. By a scaling argument [20] it may be shown that the expectation value of $:\Phi^2:$ at a point is unbounded from below as the state varies among Hadamard states. A general argument due to Epstein, Glaser and Jaffe [18] proves that loss of positivity is unavoidable for Wightman fields with vanishing vacuum expectation values.

In particular, the classical real-scalar field also satisfies the **Weak Energy Condition** (WEC): all observers measure its energy density to be nonnegative. Again, expectation values of the energy density in quantum field theory are unbounded from below at individual points. As first emphasised by Ford [30], macroscopic violations of the WEC and similar conditions could create macroscopic violations of the second law of thermodynamics. Other authors have since suggested that quantum fields might provide the negative energy densities required to support phenomena such as wormholes and warp drive. Let us see how—as Ford conjectured—quantum field theory prevents violations of the WEC at the macroscopic level. For simplicity of presentation, we actually study the Wick square, and work in Minkowski space. However, general results of this type can be proved on general globally hyperbolic spacetimes by the same means (but with more technicalities). In this sense, the following argument is a cut-down and slightly modified version of [27];[18] the core of the argument goes back to [19].

Let ω be any Hadamard state of the real Klein–Gordon field, and consider a local average

$$\omega(:\Phi^2:(f^2)) = \int d^4x\, \omega(:\Phi^2:(x)) f(x)^2$$

where f is smooth, real-valued, and supported compactly within a causally and geodesically convex neighbourhood $\mathcal{O}$. Recalling the point-splitting technique, we have

$$\begin{aligned}\omega(:\Phi^2:(f^2)) &= \int d^4x\, d^4x' \; [\omega(\Phi(x)\Phi(x') - H_{\mathcal{O}}(x,x')]\, f(x)f(x')\delta(x-x') \\ &= \int \frac{d^4k}{(2\pi)^4} \int d^4x\, d^4x' \; [\omega(\Phi(x)\Phi(x')) - H_{\mathcal{O}}(x,x')]\, f(x)f(x')e^{-ik\cdot(x-x')}\end{aligned}$$

where we have introduced a Dirac δ-function to 'unsplit' the points, and represented it as an inverse Fourier transform. Here $H_{\mathcal{O}}$ is the local Hadamard parametrix mentioned in the previous subsection; in particular, this means that the quantity in square brackets

[17] Indeed, this is true on general grounds owing to the Reeh–Schlieder theorem.

[18] Ref. [27] works with finitely many terms of the Hadamard series and uses wavefront sets sensitive to Sobolev regularity.

is symmetric under interchange of x and x'. As the integrand is unchanged if we switch x and x' and simultaneously change the sign of k, we can therefore express the integral in terms of a half-space integral, obtaining

$$\omega(:\!\Phi^2\!:(f^2)) = 2\int_{k^0>0}\frac{d^4k}{(2\pi)^4}\int d^4x\, d^4x'\, \left[\omega(\Phi(x)\Phi(x')) - H_{\mathcal{O}}(x,x')\right] f(x)f(x')e^{-ik\cdot(x-x')}.$$

Factorising the exponential and writing $f_k(x) = f(x)e^{-ik\cdot x}$, we may rewrite the right-hand side in terms of smeared quantities to find

$$\omega(:\!\Phi^2\!:(f^2)) = 2\int_{k^0>0}\frac{d^4k}{(2\pi)^4}\omega(\Phi(f_k)^*\Phi(f_k)) - 2\int_{k^0>0}\frac{d^4k}{(2\pi)^4}H_{\mathcal{O}}(\overline{f_k}, f_k),$$

where we have also used $\Phi(f_k)^* = \Phi(\overline{f_k})$. We have also assumed that the two integrals on the right-hand side exist separately, and will return to this in a moment. Proceeding with the argument, the positivity property enjoyed by states entails that the first integrand is nonnegative, and allows us to estimate

$$\omega(:\!\Phi^2\!:(f^2)) \geq -2\int_{k^0>0}\frac{d^4k}{(2\pi)^4}H_{\mathcal{O}}(\overline{f_k}, f_k), \tag{17}$$

for all Hadamard states ω. This lower bound is known as a **Quantum Inequality**; in the case of the energy density, the analogous bounds are sometimes called **Quantum Energy Inequalities**. Note that the right-hand side is independent of the state ω and depends only on the smearing function f and the local Hadamard parametrix. (For calculational purposes it is more convenient to use a partial sum of the Hadamard series rather than $H_{\mathcal{O}}$, but this introduces additional complications [27].)

However, there is an important point left to resolve: do the integrals in (6.4) exist separately? If they did not, the argument would be of little interest, as it would say only that $\omega(:\!\Phi^2\!:(f^2))$ is bounded below by $-\infty$! However, we have

$$\begin{aligned}\omega(\Phi(f_k)^*\Phi(f_k)) &= \left[(\overline{f_k}\otimes f_k)\omega_2\right]^\wedge(-k,k)\\ H_{\mathcal{O}}(\overline{f_k}, f_k) &= \left[(\overline{f_k}\otimes f_k)H_{\mathcal{O}}\right]^\wedge(-k,k);\end{aligned}$$

in other words, these are localised Fourier transforms of the type studied in the definition of the wavefront set. As we have arranged the integration region so that $k^0 > 0$, the pair $(-k,k)$ avoids the singular directions in the wavefront sets of ω_2 and $H_{\mathcal{O}}$, which are both contained in $\mathcal{N}^+\times\mathcal{N}^-$ as ω obeys the μSC. Accordingly, both integrands decay rapidly as $k\to\infty$ in the integration range, the integrals exist separately, and the lower bound is finite. This completes the derivation of the QI (17).

It should be noted that the argument used only the algebraic properties of the field and the μSC – there was no need to invoke any particular Hilbert space representation, and indeed, to do so would resulted in a loss of generality or a more complex proof.

The generalization to a QEI is straightforward, because the classical energy density can be decomposed as a sum of squares of partial differential operators applied to the field and the above argument can be applied to each such term in turn. Although the lower bound as given is somewhat involved, explicit bounds (of similar type) can be evaluated in Minkowski space, in the limiting case that f is supported on a timelike geodesic. In four dimensions, for example, this leads to the conclusion that there is no Hadamard state in which the energy density can be more negative than $-C/\tau^4$ for proper time longer than τ, where the constant C is known explicitly and is given numerically by 3.17 to three significant figures [25]. Bounds of this sort are reminiscent of the uncertainty principle [no coincidence] and restrict violations of the WEC to the microscopic scale.

Remarks:

1. There is a significant literature on Q(E)Is and their applications – see the references in [27] and the reviews [20, 21, 60].

2. QEIs do place significant constraints on the ability of quantum fields to support wormholes or other exotic spacetimes, if the fields are assumed to obey a QEI similar to those found for the free scalar fields [58, 31, 26]. The link between QEIs and thermodynamics, which originally motivated Ford [30], can be pursued abstractly (in a setting that includes the scalar field) [29].

3. The argument above, and the analogous argument for the energy density, relies on 'classical positivity' of the quantity in question. This permits a number of related bounds to be proven by similar methods, e.g., see [24] for spin-1 fields. Nonetheless, there are also QEIs for the free Dirac field [28, 13, 65] despite the fact that the 'classical' Dirac energy density is symmetrical about zero and unbounded from below. It turns out that the analogue of the Hadamard condition also functions as a microlocal version of the Dirac sea, and restores positivity [modulo a finite QEI lower bound] as well as renormalising the energy density.

4. Nonminimal coupling of the scalar field brings additional complications, because the classical field does not obey WEC, and a new type of QEI is required [23, 64]. It seems likely that something of this sort is necessary for general interacting quantum fields [56], although rigorous QEIs have been established for conformal field theories in two-dimensions [22].

7 Closing remarks and additional literature

1. To summarise: we have shown how the classical Klein–Gordon equation can be quantized (following Dirac's prescription) by both direct Hilbert space constructions and the algebraic approach. Both methods have their strengths, with the Hilbert space method particularly well-adapted (though not restricted) to situations of high symmetry, while the algebraic approach concentrates on the essential structures of the theory and treats different states in a democratic fashion. We have also seen how the concept of 'particle' becomes intrinsically ill-defined in the context of QFT in CST (or even where relatively moving observers in flat spacetime are considered); how varying gravitational fields create particles, and how Wick polynomials and the stress-energy tensor may be defined once the Hadamard class

of states is identified. Finally, we have also shown how the microlocal formulation of the Hadamard condition permits a clean specification in terms of the microlocal spectrum condition and facilitates the proof of general results such as quantum inequalities.

2. We have focussed on the construction and properties of the real linear Klein–Gordon field, and the use of the μSC to select the class of Hadamard states. There are similar constructions for the Dirac [16], Maxwell [17] and Proca [24] fields, with accompanying microlocal versions of the Hadamard condition given in [50, 39, 62] for Dirac and [24] for the spin-1 fields. See also [62] for general vector-valued fields in general spacetime dimension, and [66, 29] for a different approach to the microlocal spectrum condition in terms of distributions valued in Hilbert and Banach spaces.

3. Two important recent developments not covered in these notes, but which are closely related are: (a) the perturbative construction of interacting QFT in curved spacetimes by Brunetti & Fredenhagen [8] and Hollands & Wald [42, 43, 40]; (b) the idea of local covariance, which was key to the completion of the perturbative programme, and has been formalised in elegant mathematical terms by Brunetti, Fredenhagen and Verch [9].

References

[1] J. C. Baez, I. E. Segal, and Z. Zhou. *Introduction to Algebraic and Constructive Quantum Field Theory.* Princeton University Press, Princeton, 1992.

[2] C. Bär, N. Ginoux, and F. Pfäffle. *Wave equations on Lorentzian manifolds and quantization.* European Mathematical Society (EMS), Zürich, 2007.

[3] A. N. Bernal and M. Sánchez. Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes. *Commun. Math. Phys.*, 257:43–50, 2005, arXiv:gr-qc/0401112.

[4] A. N. Bernal and M. Sánchez. Further results on the smoothability of Cauchy hypersurfaces and Cauchy time functions. *Lett. Math. Phys.*, 77:183–197, 2006, arXiv:gr-qc/0512095.

[5] A. N. Bernal and M. Sánchez. Globally hyperbolic spacetimes can be defined as causal instead of strongly causal. *Class. Quant. Grav.*, 24:745–750, 2007, arXiv:gr-qc/0611138.

[6] N. D. Birrell and P. C. W. Davies. *Quantum fields in curved space.* Cambridge University Press, Cambridge, 1982.

[7] R. Brunetti, K. Fredenhagen, and M. Köhler. The microlocal spectrum condition and Wick polynomials of free fields on curved spacetimes. *Commun. Math. Phys.*, 180:633–652, 1996, arXiv:gr-qc/9510056.

[8] R. Brunetti and K. Fredenhagen. Microlocal analysis and interacting quantum field theories: Renormalization on physical backgrounds. *Commun. Math. Phys.*, 208:623–661, 2000, arXiv:math-ph/9903028.

[9] R. Brunetti, K. Fredenhagen, and R. Verch. The generally covariant locality principle: A new paradigm for local quantum physics. *Commun. Math. Phys.*, 237:31–68, 2003, arXiv:math-ph/0112041.

[10] D. Buchholz and J. Schlemmer. Local temperature in curved spacetime. *Class. Quant. Grav.*, 24:F25, 2007, gr-qc/0608133.

[11] L. C. B. Crispino, A. Higuchi, and G. E. A. Matsas. The Unruh effect and its applications. 2007, arXiv:0710.5373 [gr-qc].

[12] C. Dappiaggi, V. Moretti, and N. Pinamonti. Rigorous steps towards holography in asymptotically flat spacetimes. *Rev. Math. Phys.*, 18:349–416, 2006, arXiv:gr-qc/0506069.

[13] S. P. Dawson and C. J. Fewster. An explicit quantum weak energy inequality for Dirac fields in curved spacetimes. *Class. Quant. Grav.*, 23:6659–6681, 2006, arXiv:gr-qc/0604106.

[14] S. De Bièvre and M. Merkli. The Unruh effect revisited. *Class. Quant. Grav.*, 23:6525–6542, 2006, arXiv:math-ph/0604023.

[15] J. Dimock. Algebras of local observables on a manifold. *Commun. Math. Phys.*, 77:219–228, 1980.

[16] J. Dimock. Dirac quantum fields on a manifold. *Trans. Am. Math. Soc.*, 269:133–147, 1982.

[17] J. Dimock. Quantized electromagnetic field on a manifold. *Rev. Math. Phys.*, 4:223–233, 1992.

[18] H. Epstein, V. Glaser, and A. Jaffe. Nonpositivity of the energy density in quantized field theories. *Il Nuovo Cim.*, 36:1016–1022, 1965.

[19] C. J. Fewster. A general worldline quantum inequality. *Class. Quant. Grav.*, 17:1897–1911, 2000, arXiv:gr-qc/9910060.

[20] C. J. Fewster. Energy inequalities in quantum field theory. In J. C. Zambrini, editor, *XIVth International Congress on Mathematical Physics*, Singapore, 2005. World Scientific. An expanded and updated version is available as arXiv:math-ph/0501073.

[21] C. J. Fewster. Quantum energy inequalities and stability conditions in quantum field theory. In A. Boutet de Monvel, D. Buchholz, D. Iagolnitzer, and U. Moschella, editors, *Rigorous Quantum Field Theory: A Festschrift for Jacques Bros*, volume 251 of *Progress in Mathematics*, Boston, 2006. Birkhäuser, arXiv:math-ph/0502002.

[22] C. J. Fewster and S. Hollands. Quantum energy inequalities in two-dimensional conformal field theory. *Rev. Math. Phys.*, 17:577, 2005, arXiv:math-ph/0412028.

[23] C. J. Fewster and L. W. Osterbrink. Quantum energy inequalities for the non-minimally coupled scalar field. *J. Phys.*, A41:025402, 2008, arXiv:0708.2450 [gr-qc].

[24] C. J. Fewster and M. J. Pfenning. A quantum weak energy inequality for spin-one fields in curved spacetime. *J. Math. Phys.*, 44:4480–4513, 2003, arXiv:gr-qc/0303106.

[25] C. J. Fewster and M. J. Pfenning. Quantum energy inequalities and local covariance. I: Globally hyperbolic spacetimes. *J. Math. Phys.*, 47:082303, 2006, arXiv:math-ph/0602042.

[26] C. J. Fewster and T. A. Roman. On wormholes with arbitrarily small quantities of exotic matter. *Phys. Rev.*, D72:044023, 2005, arXiv:gr-qc/0507013.

[27] C. J. Fewster and C. J. Smith. Absolute quantum energy inequalities in curved spacetime. *Annales Henri Poincare*, 9:425–455, 2008, arXiv:gr-qc/0702056.

[28] C. J. Fewster and R. Verch. A quantum weak energy inequality for Dirac fields in curved spacetime. *Commun. Math. Phys.*, 225:331–359, 2002, arXiv:math-ph/0105027.

[29] C. J. Fewster and R. Verch. Stability of quantum systems at three scales: Passivity, quantum weak energy inequalities and the microlocal spectrum condition. *Commun. Math. Phys.*, 240:329–375, 2003, arXiv:math-ph/0203010.

[30] L. H. Ford. Quantum coherence effects and the second law of thermodynamics. *Proc. Roy. Soc. Lond. A*, 364:227–236, 1978.

[31] L. H. Ford and T. A. Roman. Quantum field theory constrains traversable wormhole geometries. *Phys. Rev.*, D53:5496–5507, 1996, arXiv:gr-qc/9510071.

[32] S. A. Fulling, M. Sweeny, and R. M. Wald. Singularity structure of the two point function in quantum field theory in curved space-time. *Commun. Math. Phys.*, 63:257–264, 1978.

[33] S. A. Fulling. *Aspects of Quantum Field Theory in Curved Space-Time*. Cambridge University Press, Cambridge, 1989.

[34] S. A. Fulling and S. N. M. Ruijsenaars. Temperature, periodicity and horizons. *Phys. Rept.*, 152:135–176, 1987.

[35] G. W. Gibbons and S. W. Hawking. Cosmological event horizons, thermodynamics, and particle creation. *Phys. Rev.*, D15:2738–2751, 1977.

[36] R. Haag. *Local Quantum Physics: Fields, Particles, Algebras*. Springer-Verlag, Berlin, 1992.

[37] S. W. Hawking. Particle creation by black holes. *Commun. Math. Phys.*, 43:199–220, 1975.

[38] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge University Press, London, 1973.

[39] S. Hollands. The Hadamard condition for Dirac fields and adiabatic states on Robertson-Walker spacetimes. *Commun. Math. Phys.*, 216:635–661, 2001, arXiv:gr-qc/9906076.

[40] S. Hollands. Renormalized quantum Yang-Mills fields in curved spacetime. 2007, arXiv:0705.3340 [gr-qc].

[41] S. Hollands and W. Ruan. The state space of perturbative quantum field theory in curved spacetimes. *Annales Henri Poincaré*, 3:635–657, 2002.

[42] S. Hollands and R. M. Wald. Local Wick polynomials and time ordered products of quantum fields in curved spacetime. *Commun. Math. Phys.*, 223:289–326, 2001, arXiv:gr-qc/0103074.

[43] S. Hollands and R. M. Wald. Existence of local covariant time ordered products of quantum fields in curved spacetime. *Commun. Math. Phys.*, 231:309–345, 2002, arXiv:gr-qc/0111108.

[44] L. Hörmander. *The Analysis of Linear Partial Differential Operators I.* Springer Verlag, Berlin, 1983.

[45] W. Junker. Hadamard states, adiabatic vacua and the construction of physical states for scalar quantum fields on curved space- time. *Rev. Math. Phys.*, 8:1091–1159, 1996.

[46] W. Junker and E. Schrohe. Adiabatic vacuum states on general spacetime manifolds: Definition, construction, and physical properties. *Annales Poincare Phys. Theor.*, 3:1113–1182, 2002, arXiv:math-ph/0109010.

[47] B. S. Kay. Quantum field theory in curved space-time. In K. Schmüdgen, editor, *Mathematical Physics X: Proceedings of the Xth Congress on Mathematical Physics*, pages 383–387, Berlin, 1992. Springer-Verlag.

[48] B. S. Kay. Quantum field theory in curved spacetime. In J.-P. Françoise, G. L. Naber, and S. T. Tsou, editors, *Encyclopedia of Mathematical Physics*, volume 4, pages 202–214. Academic Press/Elsevier Science, Oxford, 2006, arXiv:gr-qc/0601008.

[49] B. S. Kay and R. M. Wald. Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on space-times with a bifurcate killing horizon. *Phys. Rept.*, 207:49–136, 1991.

[50] K. Kratzert. Singularity structure of the two point function of the free Dirac field on a globally hyperbolic spacetime. *Annalen Phys.*, 9:475–498, 2000, arXiv:math-ph/0003015.

[51] S.-Y. Lin and B. L. Hu. Backreaction and the Unruh effect: New insights from exact solutions of uniformly accelerated detectors. *Phys. Rev.*, D76:064008, 2007, arXiv:gr-qc/0611062.

[52] J. Louko and A. Satz. Transition rate of the Unruh-DeWitt detector in curved spacetime. *Class. Quant. Grav.*, 25:055012, 2008, arXiv:0710.5671 [gr-qc].

[53] V. Moretti. Quantum ground states holographically induced by asymptotic flatness: Invariance under spacetime symmetries, energy positivity and Hadamard property. *Commun. Math. Phys.*, 279:31–75, 2008, arXiv:gr-qc/0610143.

[54] V. Moretti. Comments on the stress-energy tensor operator in curved spacetime. *Commun. Math. Phys.*, 232:189–221, 2003, arXiv:gr-qc/0109048.

[55] H. Olbermann. States of low energy on Robertson-Walker spacetimes. *Class. Quant. Grav.*, 24:5011–5030, 2007, arXiv:0704.2986 [gr-qc].

[56] K. D. Olum and N. Graham. Static negative energies near a domain wall. *Phys. Lett.*, B554:175–179, 2003, arXiv:gr-qc/0205134.

[57] B. O'Neill. *Semi-Riemannian Geometry.* Academic Press, New York, 1983.

[58] M. J. Pfenning and L. H. Ford. The unphysical nature of 'warp drive'. *Class. Quant. Grav.*, 14:1743–1751, 1997, arXiv:gr-qc/9702026.

[59] M. J. Radzikowski. Micro-local approach to the Hadamard condition in quantum field theory on curved space-time. *Commun. Math. Phys.*, 179:529–553, 1996.

[60] T. A. Roman. Some thoughts on energy conditions and wormholes. In S. Bergliaffa and M. Novello, editors, *Proceedings of the Tenth Marcel Grossmann Conference on General Relativity*, Singapore, 2006. World Scientific, gr-qc/0409090.

[61] H. Sahlmann and R. Verch. Passivity and microlocal spectrum condition. *Commun. Math. Phys.*, 214:705–731, 2000, arXiv:math-ph/0002021.

[62] H. Sahlmann and R. Verch. Microlocal spectrum condition and Hadamard form for vector valued quantum fields in curved space-time. *Rev. Math. Phys.*, 13:1203–1246, 2001, arXiv:math-ph/0008029.

[63] J. A. Sanders. *Aspects of locally covariant quantum field theory.* PhD thesis, University of York, 2008, arXiv:0809.4828 [math-ph].

[64] J. Schlemmer and R. Verch. Local thermal equilibrium states and quantum energy inequalities. *Annales Henri Poincare*, 9:945–978, 2008, arXiv:0802.2151 [gr-qc].

[65] C. J. Smith. An absolute quantum energy inequality for the Dirac field in curved spacetime. *Class. Quant. Grav.*, 24:4733–4750, 2007, arXiv:0705.2203 [gr-qc].

[66] A. Strohmaier, R. Verch, and M. Wollenberg. Microlocal analysis of quantum fields on curved spacetimes: Analytic wavefront sets and Reeh-Schlieder theorems. *J. Math. Phys.*, 43:5514–5530, 2002, arXiv:math-ph/0202003.

[67] W. G. Unruh. Notes on black hole evaporation. *Phys. Rev.*, D14:870–892, 1976.

[68] R. M. Wald. The back reaction effect in particle creation in curved space-time. *Commun. Math. Phys.*, 54:1–19, 1977.

[69] R. M. Wald. Trace anomaly of a conformally invariant quantum field in curved space-time. *Phys. Rev.*, D17:1477–1484, 1978.

[70] R. M. Wald. *General Relativity.* University of Chicago Press, Chicago, 1984.

[71] R. M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* University of Chicago Press, Chicago, 1994.