What is a Photon?

Quantum Censorship, Blackbody Catastrophe, and Superfluid He$^4$

Frank Wilczek
Center for Theoretical Physics, MIT, Cambridge MA 02139 USA

January 25, 2014

Abstract

A photon, according to modern quantum field theory, is a minimal change in the wave function for the electromagnetic field or, less formally, in the probability distribution for what detectors of electromagnetic radiation will sense. Here I explain that strange formulation, by building up to it through the general concept of quantum censorship. From this perspective, the resolution of blackbody radiation catastrophe has much in common with Feynman’s theory of superfluid $^4$He.

The resolution, in quantum theory, of the ultraviolet catastrophe of black-body radiation theory can be viewed in various ways. Here I discuss it as a manifestation of quantum censorship – that is, lifting of energy degeneracy among classical modes through their re-organization into coherent superpositions.

The difficulty of classical physics that initiated the quantum revolution was its (catastrophic) prediction of infinite energy for blackbody radiation at any non-zero temperature. The catastrophe arises because of the great abundance of high-frequency modes. In thermal equilibrium the average amplitudes for these modes are small, individually, but add up to infinite energy, contrary to observation. It is interesting to revisit that problem, in those concrete terms, in the context of modern, mature quantum theory. Of course, there is no question of deriving essentially new results for this old problem, but it can be entertaining to view things from different points of view, and it may turn out that doing so develops intuition that will prove
useful in other contexts. That philosophy was championed by Feynman, both as methodological advice and in his practice.

To elucidate of quantum censorship for electromagnetic fields we will need to consider its quantum mechanical wave function. Since much of the technology of modern quantum field theory was developed precisely to avoid the necessity of considering wave functions, this approach, although “elementary” and logical, may appear strange and unfamiliar even to experts. Accordingly I will build up the necessary concepts in two stages, first elaborating the meaning of quantum censorship in toy problems of one-particle quantum mechanics, and then generalizing to field theory.

I will adopt units such that $\hbar = c = k = 1$ throughout.

**Elementary Quantum Censorship Examples**

Let us begin by considering a free particle on a ring, with its position parameterized by an angle $\phi$. The Hamiltonian is

$$H = \frac{p^2_{\phi}}{2mR^2} = \frac{p^2_{\phi}}{2I} \quad (1)$$

where $p_{\phi}$ is the momentum conjugate to $\phi$, $m$ is the mass of the particle, $R$ is the radius of the ring, and $I = mR^2$ is the moment of inertia.

Classically, the minimum energy states form a continuum, corresponding to the particle at rest ($p_{\phi} = 0$) at any angle $\phi$.

Quantum mechanically, states are specified as superpositions of the position eigenstates $|\phi\rangle$. In Dirac notation, they take the form

$$|\psi\rangle = \frac{2\pi}{\int_0^{2\pi} \psi(\phi) |\phi\rangle} \quad (2)$$

which we commonly write simply as the wave function $\psi(\phi)$. The lowest energy state is unique, specified by the constant (normalized) function

$$\psi_0(\phi) = \frac{1}{\sqrt{2\pi}} \quad (3)$$

Its energy $E_0 = 0$. The next-highest energy states are

$$\psi_{\pm1}(\phi) = \frac{1}{\sqrt{2\pi}} e^{\pm i\phi} \quad (4)$$

---

1Particle quantum mechanics itself can be considered as quantum field theory is 0 space, 1 time dimension.
with energy $E_{\pm 1} = \frac{1}{I}$. Thus there is a finite energy gap, separating a unique lowest-energy state from any other state.

The general eigenstates and their energies are

$$
\psi_l(\phi) = \frac{1}{\sqrt{2\pi}} e^{il\phi}
$$

$$
E_l = \frac{l^2}{2I}
$$

(5)

where $l$ runs over the integers. By forming superpositions, of these states, we can recover the positional states $|\phi\rangle$, according to the inverse Fourier transformation

$$
|\phi\rangle = \frac{1}{\sqrt{2\pi}} \sum_l e^{-il\phi} \psi_l(\phi)
$$

(6)

Thus the position variable has not been irrevocably lost. There is still a continuum of position states, though they are no longer low energy states. (Indeed, their energy is infinite.)

This simple example demonstrates, in a transparent way, the essential features of quantum censorship. In place of the vast number of low-energy states we found in the classical theory, we find a unique ground state, separated by an energy gap from all excited states. If we probe the system at small energy (compared to the gap) we will have no effective degrees of freedom, though the effective “vacuum” may exhibit non-trivial response. Or, if we choose to view our system from afar – as a particle, as opposed to a universe – we would say that the particle exhibits no internal degrees of freedom. It reacts, but does not change, in response to small perturbations. It thereby mimics the “hard, massy, impenetrable” objects of classical atomism.

Our simple model comes up in the physical context of molecular physics. In the Born-Oppenheimer approximation, we first derive a “semi-classical” picture of the molecule with nuclear positions fixed, and then allow for nuclear motion. The simplest such motion is overall rotation, and our toy model describes the formation of rotational bands, with $\phi$ as the orientation angle and $mR^2$ the moment of inertia. (Strictly speaking we must allow for rotations in three dimensions, but the essential logic is similar.) In that context quantum censorship explains the absence of a true electric dipole moment at ultra-low electric fields. Instead we find a quadratic response to low-energy electric fields, as required by $P$ and $T$ symmetry. When the energy associated with the electric field interaction greatly exceeds the energy gap, the expected semi-classical behavior re-emerges, and we find
the effective molecular electric dipole moments reported in chemical tables. Quantum censorship is not absolute; it can be overcome by resolute probes of the right sort.

Now let us consider another sort of probe – exposing the system to a temperature bath, bringing it into equilibrium at temperature $T$. The most obvious question, which also gets to the heart of the blackbody catastrophe, is: How much energy does that require?

Classically, the energy at temperature $T$ is

$$\mathcal{E} = \frac{\int dp\phi \frac{p^2}{2T} e^{-\frac{p^2}{2T}}}{\int dp\phi e^{-\frac{p^2}{2T}}} = \frac{T}{2} \quad (7)$$

embodying the classical equipartition theorem.

Quantum mechanically the result is different, especially at low temperatures. The energy is a sum over a discrete spectrum of states, rather than an integral:

$$\mathcal{E} = \frac{\sum l^2 \frac{1}{2\pi} e^{-\frac{l^2}{2\pi T}}}{\sum l e^{-\frac{l^2}{2\pi T}}} \quad (8)$$

When $T \gg I^{-1}$, it is a good approximation to we can replace the sums by integrals, and we get back to the classical result. But when $T \ll I^{-1}$ the exponential factors become very small, and we have approximately

$$\mathcal{E} \xrightarrow{T \to 0} \frac{1}{I} e^{-\frac{1}{2\pi T}} \quad (9)$$

Here we find that quantum censorship is exponentially effective. In this sense, heat baths are very soft probes. Since they do not couple to the underlying local degrees of freedom directly, they do not much disturb the Quantum Censor.

The preceding discussion applies, in its qualitative essentials, to the slightly more complicated case of the harmonic oscillator, which turns out to be important for field theory. I will record the parallel steps with minimal commentary. The Hamiltonian is

$$H = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2 \quad (10)$$

The classical theory allows a continuum of low-energy states, labelled by $(p, x)$ with small $p$ and $x$. In the quantum theory there is still a continuum
of states labelled by position, as $|x\rangle$, but their energy is infinite. More useful are the states of definite energy. They are labelled with a discrete index $n$, where $n \geq 0$ is an integer. The energies are

$$E_n = (n + \frac{1}{2}) \omega$$  \hspace{1cm} (11)

There is an energy gap $\omega$ between the unique ground state and any excited states. The corresponding wave functions can be found in textbooks; they involve Hermite polynomials modulated by Gaussian factors $e^{-mωx^2}$. For later reference, the first two are

$$\psi_0(x) = \left(\frac{mω}{\pi}\right)^{\frac{1}{4}} e^{-\frac{mωx^2}{2}}$$ \hspace{1cm} (12)

$$\psi_1(x) = \left(\frac{mω}{\pi}\right)^{\frac{1}{4}} \sqrt{mωx} e^{-\frac{mωx^2}{2}}$$ \hspace{1cm} (13)

Classically, thermal energy is

$$E = T$$ \hspace{1cm} (14)

while the quantum thermal energy is

$$E = \sum_{n=0}^{\infty} (n + \frac{1}{2}) \omega e^{-\frac{(n + \frac{1}{2})ω}{T}} \sum_{n=0}^{\infty} e^{-\frac{(n + \frac{1}{2})ω}{T}} = \frac{ω e^{-\frac{T}{2}}}{1 - e^{-\frac{ω}{T}}}$$ \hspace{1cm} (15)

As for the rotator, we find the quantum energy approaches the classical value for $T \gg ω$, but is exponentially small, instead of linear, for $T \ll ω$.

A notable difference between the rotator and the harmonic oscillator is that the oscillator, but not the rotator, exhibits zero-point energy. Physically, this occurs because the delocalizing effect of gradient energy, embodied in the Hamiltonian after the substitution $p \rightarrow -i \frac{d}{dx}$, can be accommodated by having a perfectly uniform wave function, without any price in potential energy. If we had considered a free particle on an interval, instead of a circle, we would have found non-trivial zero point energy. The zero point energy is not directly observable, since it rests on a comparison of different models of physical behavior (classical versus quantum) as opposed to different phenomena within a single theory. But it is significant, reflecting that the ground state wave function assigns non-zero amplitude to position states that do not minimize the potential energy.
The Wave Functional in Quantum Field Theory: Concept

In quantum field theory proper the dynamical variables are fields (i.e., functions) $\phi^j(x)$, as opposed to simply positions $\phi$ or $x$. (We allow for an “internal space” index $j$.) The states are specified as superpositions of the configurations of these variables. Thus they are functionals (functions of functions):

$$\Psi = \int D\phi^j(x) \Psi(\phi^j(x))|\phi^j(x)\rangle$$

(16)

As before, we also denote this state simply by the wave functional $\Psi(\phi^j(x))$.

We will work up to the electromagnetic field by first considering a real massive scalar field $\phi(x)$, with no internal space labels. The Hamiltonian is

$$H = \int dx \left( \frac{1}{2} \Pi(x)^2 + \frac{1}{2} \partial^2 + \frac{M^2}{2} \phi^2 \right)$$

(17)

where $\Pi(x)$ is the dynamical variable canonically conjugate to $\phi(x)$. It is useful to analyze $\phi(x)$ into spatial eigenmodes $\phi_n(x)$, that satisfy

$$(-\partial^2 + M^2) \phi_n(x) = \omega_n^2 \phi_n(x)$$

(18)

and appropriate boundary conditions. A standard choice, adequate for our purposes, is to consider fields that are periodic, with period $L$, in all three directions. Then we have a complete set of eigenmodes labelled by spatial frequencies $(k_1, k_2, k_3)$ subject to

$$k_j = \frac{2\pi n_j}{L}$$

(19)

where the $n_j$ are integers, in the form

$$\phi_{(k_1,k_2,k_3)}(x) = \left(\frac{1}{L}\right)^{\frac{3}{2}} e^{i(k_1x_1+k_2x_2+k_3x_3)}$$

(20)

(Strictly speaking, we should treat the real and imaginary parts of this solution separately. Thus there are two modes for a given $\vec{k}$, but the modes for $\pm \vec{k}$ are the same. I will not belabor this sort of thing, but just suppose that one assigns the cosine solution to one of $\pm \vec{k}$ and the sine to the other, so we can use real coefficients in our expansion, yet have a simple sum over $\vec{k}$.) These are eigenmodes with

$$\omega_{(k_1,k_2,k_3)}^2 = k_1^2 + k_2^2 + k_3^2 + M^2$$

(21)
We also have a corresponding complete set of modes for $\Pi(x)$.

Now when we expand field an arbitrary field (or conjugate field) in terms of eigenmodes we have

$$
\phi(x) = \sum_{(k_1,k_2,k_3)} c_{(k_1,k_2,k_3)} \phi_{(k_1,k_2,k_3)}(x)
$$

$$
\Pi(x) = \sum_{(k_1,k_2,k_3)} d_{(k_1,k_2,k_3)} \Pi_{(k_1,k_2,k_3)}(x)
$$

(22)

The classical state $(\Pi(x), \phi(x))$ has energy

$$
H(\Pi(x), \phi(x)) = \sum_{(k_1,k_2,k_3)} \frac{1}{2} (c_k^2 \vec{k}^2 + M^2) + d_k^2 \vec{k}^2)
$$

(23)

If we regard $(k_1, k_2, k_3)$ as a label, and $c_k, d_k$ as values of a effective positions and momenta nominally attached to that label, then this is the same Hamiltonian we’d have for an harmonic oscillator with mass $m = 1$ and (frequency)$^2 \omega^2 = \vec{k}^2 + M^2$. Thus our field represents, in effect, an infinite number of independent harmonic oscillators.

We can, therefore, easily apply the results of our preceding analysis to this new context.

Classically, there are an enormous number of low-energy states. In particular, one has many low-energy states with contributions from large $\vec{k}$ and $M$, since the values of $c_k$ and $d_k$ can be small.

Classically, we should assign energy $T$ to each oscillator. Since there are infinitely many of them, that is physically catastrophic – one cannot reach equilibrium, at any finite temperature\(^2\). This is, of course, the essence of the blackbody radiation catastrophe.

In quantum mechanics, we eventually arrive at harmonic oscillators too, but the path is steeper – and, importantly, leads us to elucidate exactly what it is that “oscillates”. The states are specified by field configurations, and the fields become operators. (To disambiguate the notation we put a hat over operators, writing e. g. $\hat{\phi}(x)$ for the operator associated with the original classical field variable $\phi(x)$.) In this state space the field operator acts as multiplication, so

$$
\hat{\phi}(x)|\phi\rangle = \phi(x)|\phi\rangle
$$

(24)

\(^2\)Strictly speaking, it is not logically catastrophic – the possibility of thermal equilibrium is an empirical observation, not an a priori requirement.
The canonical conjugate field \( \hat{\Pi}(x) \) is realized as the functional derivative operation
\[
\hat{\Pi}(x) = -i \frac{\delta}{\delta \phi(x)}
\] (25)

Putting these together, we get the stationary Schrödinger equation for \( \Psi(\phi) \), in the form
\[
E \Psi(\phi(x)) = \frac{1}{2} \int dy \left( -\left( \frac{\delta}{\delta \phi(y)} \right)^2 + (\partial \phi(y))^2 + M^2 \phi^2(y) \right) \Psi(\phi(x))
\] (26)

The functional Laplacian (and especially its spatial integral), which appears here as the first term on the right-hand side, assumes a simple form in the coordinates \( c_{\vec{k}} \), as we now demonstrate. We have
\[
\frac{\delta}{\delta \phi(y)} = \sum_{\vec{k}} \frac{\delta c_{\vec{k}}}{\delta \phi(y)} \frac{\delta}{\delta c_{\vec{k}}}
\] (27)
so the first step is to calculate the \( \frac{\delta c_{\vec{k}}}{\delta \phi(y)} \). We do that by considering the effect of incrementing \( \phi \) by a Dirac \( \delta \) function, according to
\[
\frac{\delta \phi(x)}{\delta \phi(y)} = \sum_{\vec{k}} \frac{\delta c_{\vec{k}}}{\delta \phi(y)} \phi_{\vec{k}}(x) = \delta(x - y)
\] (28)

Multiplying the two latter expressions by \( \int dx \phi_{\vec{k}}(x) \) and integrating, we find
\[
\frac{\delta c_{\vec{k}}}{\delta \phi(y)} = \phi_{\vec{k}}(y)
\] (29)
Thus
\[
\left( \frac{\delta}{\delta \phi(y)} \right)^2 = \sum_{\vec{k}} \frac{\delta c_{\vec{k}}}{\delta \phi(y)} \frac{\delta}{\delta c_{\vec{k}}} \phi_{\vec{k}}(y) \phi_{\vec{k}}(y)
\] (30)
And now integrating over \( y \), we find, exploiting the orthonormality of the \( \phi_{\vec{k}} \),
\[
\int dy \left( \frac{\delta}{\delta \phi(y)} \right)^2 = \sum_{\vec{k}} \left( \frac{\delta c_{\vec{k}}}{\delta \phi(y)} \right)^2
\] (31)

The Wave Functional in Quantum Field Theory: Result and Consequences

With that realization, our Schrödinger equation Eqn. (26) becomes
\[
E \Psi(\{ c_{\vec{k}} \}) = \frac{1}{2} \left( -\sum_{\vec{k}} \left( \frac{\delta}{\delta c_{\vec{k}}} \right)^2 + (\omega^2_{\vec{k}} + M^2) c_{\vec{k}}^2 \right) \Psi(\{ c_{\vec{k}} \})
\] (32)
This Equation (32) is a major result\(^3\). It allows us to find the ground state, and all the excited states, in simple, explicit forms. Indeed, in this representation the Hamiltonian (right hand side) is a sum of independent terms, one for each allowed momentum \(\vec{k}\), each of which is simply an harmonic oscillator. So we can find a complete set of energy eigenstates in the factorized form

\[ \Psi(\{c_k\}) = \prod_k \psi_k(c_k) \tag{33} \]

where each \(\psi_k\) is an eigenstate for a harmonic oscillator with (frequency)\(^2\)

\[ \omega_k^2 = \vec{k}^2 + M^2 \]

(and mass \(m = 1\)). The energy for the total state \(\Psi\) will be the sum of the energies for all these separate single-mode harmonic oscillators.

Note that what “oscillates”, in these single-mode oscillators, is the amplitude of the corresponding field mode. In the overall ground state, each of those amplitudes will participate in its own individual harmonic oscillator ground state. In particular, we have the phenomenon of zero-point “motion”: There is a non-trivial (Gaussian) probability distribution for each amplitude, and they are independent.

The form of the ground state and of the excitations for the (free) electromagnetic field are, in essence, the same as what we’ve just worked out for a scalar field. One has \(M = 0\), and a doubling reflecting the two possible polarizations of transverse field vibration modes, but otherwise the basic structure, with independent harmonic oscillators for each mode, is identical. I’ll spell out the details momentarily. Anticipating that result, I’d like to draw out a few consequences.

1. Space, as empty as we can make it, will be found to contain non-zero fields, whenever we make field-dependent observations. And here I mean “we” very broadly, to include for instance our wee brother electrons. They are ever in touch with a sea of active oscillators, and it most definitely affects their properties.

2. Formally, the total energy of ground state diverges, due to an accumulation of zero-point energies. As discussed previously, however, that notional energy is not directly physical, since it is defined relative to a different theory – the classical theory – and not intrinsically within the quantum theory. Within the quantum theory, the energy of the ground state can be taken to define the zero of energy.

\(^3\)Though hardly a new one, of course.
3. For gravity the absolute value of the energy matters, not only relative values. A universal ground-state energy, in a relativistic theory, will contribute to the equations of general relativity in same way as Einstein’s cosmological term (rechristened dark energy in recent literature). One can regard its value as part of the definition of the theory of gravity, and if one takes that attitude then again the diverging total zero-point energy has no independent significance. Nevertheless, many physicists find it disturbing. They feel that zero point energy is a genuine physical phenomenon. It reflects the same spontaneous activity of quantum fields, other of whose consequences have been manifested in important, observable (and actually observed!) effects including vacuum polarization in quantum electrodynamics and asymptotic freedom in quantum chromodynamics. Thus they feel that the observed finite, and (by most standards) very small value of the cosmological term must also additional contributions, that cancel the zero point energy of electromagnetic fields. In supersymmetric theories there are cancellations between boson and fermion zero point energies. This might be taken as a philosophical point in favor of supersymmetry.

4. For modes at any finite frequency $\omega$, there is an energy gap $\omega$. There is no possibility of small amplitude, high frequency but low energy, oscillations. Thus quantum censorship applies, in this very concrete way, to the electromagnetic field.

5. One is accustomed to the idea that characteristically quantum effects only become evident at low temperatures. But low compared to what? Clearly, what is most relevant is the ratio of temperature to the energy gaps. For any finite temperature $T$, there are modes with $T \sim \omega$, and also modes with $T \ll \omega$. In this sense, for the electromagnetic field all temperatures are low temperatures. This circumstance helps to explain why quantum phenomena were first distinctly encountered, historically, in the blackbody radiation problem.

6. There is a profound, instructive complementarity between the two results, that the field contains considerable spontaneous activity, and that its ground state is unique and (for any finite frequency mode) separated by an energy gap from any excitation. The crux is that the pattern of spontaneous activity is self-generated, and has a kind of cohesion. Small perturbations will change the pattern slightly, but enough into bring in the qualitatively different patterns characteristic of excited states.
7. Although the dynamical equations are strictly local, their low-energy solutions are highly structured and, at the lowest energies, extended in space. For large $\vec{k}$ there are many excitations with nearly the same energy and wavevector, and by superposing them we can make wavepacket modes, that are more nearly, but still imperfectly, localized in space. This genuine form of non-locality, introduced into solutions of quantum problems by restriction of energy, is distinct from the much-discussed pseudo-nonlocality of quantum entanglement situations.

Let me now return to the electromagnetic field proper, as opposed to the scalar field treated above. I will be very brief, because the relevant mathematics is treated in many standard texts. The modern approach is to start from the Maxwell Lagrangian density

$$L = \frac{1}{2} \vec{E}^2 - \frac{1}{2} \vec{B}^2 = \frac{1}{2} (\partial_0 \vec{A} - \vec{\partial} A_0)^2 - \frac{1}{2} (\vec{\partial} \times \vec{A})^2$$

expressed in terms of potentials. $A_0$ occurs without a time derivative, so its variation yields the Gauss' law $\vec{\partial} \cdot \vec{E} = 0$ as a constraint. Having imposed the constraint, we can choose $A_0 = 0$ as a gauge condition. Then Gauss law constraint $\vec{\partial} \cdot \partial_0 \vec{A} = 0$ because a condition

$$\vec{\partial} \cdot \vec{\Pi} = 0$$

on the canonical momenta

$$\vec{\Pi} = \frac{\delta L}{\delta \partial_0 \vec{A}}$$

This eliminates the longitudinal canonical momentum. The longitudinal vector potential itself likewise does not appear, since it does not contribute to the curl. So one has only transverse degrees of freedom. When we expand those in modes, the Hamiltonian reduces to two copies of the scalar field Hamiltonian, with $M = 0$ and with the internal space indices running over two directions perpendicular to $\vec{k}$, for the two transverse directions of $\vec{A}_{\vec{k}}$.

And now we can finally discuss the question of the title: “What is a photon?” concretely, in terms of fields in space-time. The excited states of the electromagnetic field correspond to the excited states of the independent oscillators we have analyzed it into. Thus we obtain the wave functional of a photon with three-momentum $\vec{k}$ by putting the corresponding oscillator into its first excited state. That is the answer to our question.
From a space-time point of view, in terms of the field variables, a photon is evidently a pretty complicated object. Most of its complex structure arises from the complexity of the ground state, i.e. “vacuum” or “empty space”. What’s more, I’ve been ignoring the effect of interactions between the electromagnetic fields and other fields. When we include the effects of interactions, both the ground state and the excitations get much more intricate.

Quantum field theorists have developed clever tricks to factor out the ground state in many kinds of calculations. The most vivid and versatile example is the Feynman graph technology for doing calculations of scattering processes in perturbation theory. In that technique, photons appear as very simple objects and the ground state as nothing at all. But in many important cases no such trick is available. In the lattice gauge theory approach to quantum chromodynamics, one constructs the ground state directly, as a functional in space-time. Doing a decent job of that pushes the limits of modern computer power, both for computation – and for memory to store the result! The excitations – the objects we call pions, protons, and so forth – are discovered by adding disturbances to the ground state, and studying what they settle into. Both the ground state and the excitations are many steps removed, both conceptually and materially, from the underlying quark and gluon fields used to formulate the theory. Perhaps someone will discover some new tricks, and the straightforward, “elementary” or “brute force” approach of lattice gauge theory\(^4\) will be streamlined, or bypassed. So far that hasn’t happened, despite many years of hard efforts by brilliant people, and there are serious reasons to think it never will, at least for accurate work. Indeed, since lattice gauge theory has supplied us with accurate answers, which reproduce Nature, we can look at those answers, and perceive their richness and complexity. It strains credulity that such complex answers could be generated from simple rules, without the intervention of very substantial computations.

**Superfluid \(^4\)He**

[This part is under construction. Its inspiring thought is that the essence of Feynman’s theory of the superfluidity of helium 4 is to argue away many apparent low-energy excitations, which if present could cause dissipation. This is quite similar to how we argue away apparent low-energy excitations in classical electromagnetic fields, when we pass to the quantum theory.]

\(^4\)See also: Shut up and calculate!
That thought can be made considerably more precise.]