$E=mc^2$

Mitchell J. Feigenbaum, and N. David Mermin

Citation: American Journal of Physics **56**, 18 (1988); doi: 10.1119/1.15422
View online: https://doi.org/10.1119/1.15422
View Table of Contents: https://aapt.scitation.org/toc/ajp/56/1
Published by the American Association of Physics Teachers

$E=mc^2$

Mitchell J. Feigenbaum and N. David Mermin
Laboratory of Atomic and Solid Physics, Cornell University, Ithaca, New York 14853-2501

(Received 2 March 1987; accepted for publication 20 April 1987)

A purely mechanical version is given of Einstein's 1905 argument that the mass of a body depends on its energy content. The *Gedankenexperiment* described here is the same as Einstein's, except that the body loses energy not through electromagnetic radiation, but through the emission of massive particles. The concept of mass is not assumed, but is extracted as one of two constants of integration. Viewing mass in this way emphasizes a sometimes obscured distinction between mass as a proportionality constant in the kinetic energy (for which $E=mc^2$ has profound physical content), and mass as it appears in the conventional relativistic rest energy (for which $E=mc^2$ is merely a convenient convention).

The first argument relating the mass of a body to its energy content was given by Einstein in 1905.[1] In Einstein's *Gedankenexperiment* the change in energy content is produced by the emission of light and his analysis relies upon the transformation law for the energy of electromagnetic radiation. Since 1905 a variety of different arguments have been constructed deriving the relation without appealing to electrodynamics.[2] We would like to describe here a derivation that, while purely mechanical, extracts $E=mc^2$ by directly considering a single inelastic process very much like the one Einstein used in 1905. Our argument also brings out the important but sometimes obscured distinction between mass, the proportionality constant in the kinetic energy, on which $E=mc^2$ has a direct and profound bearing, and mass, the rest energy divided by c^2, for which $E=mc^2$ is little more than a convenient convention.

We first review Einstein's 1905 argument. Consider a particle at rest with energy content (*Energieinhalt*) E_1, that emits in opposite directions two identical quantities of light (*Lichtmenge*), each with energy ϵ. After the emission the particle remains at rest with energy content E_2. Energy conservation requires that

$$E_1=E_2+2\epsilon. \tag{1}$$

Consider next the same decay in a frame of reference moving with velocity v along a direction making an angle θ with the direction of propagation of the emitted light. Call the initial and final energies of the particle in that frame $E_1(v)$ and $E_2(v)$. Einstein knew from electrodynamics how the energies of the two quantities of light transformed, and therefore he knew that energy conservation in the moving frame required

$$E_1(v)=E_2(v)+\frac{\epsilon(1-v\cos\theta/c)}{\sqrt{1-v^2/c^2}}+\frac{\epsilon(1+v\cos\theta/c)}{\sqrt{1-v^2/c^2}}. \tag{2}$$

Subtraction of (1) from (2) gives

$$[E_1(v)-E_1]-[E_2(v)-E_2]=2\epsilon(1/\sqrt{1-v^2/c^2}-1). \tag{3}$$

But $E(v)-E$ (with either subscript) is just the difference in the energy of a particle moving with speed v and the energy of the same particle when at rest, i.e., it is the kinetic energy of the particle $K(v)$ which, when v is small compared to c, is just $\frac{1}{2}mv^2$. In this same limit the right-hand side of Eq. (3) simplifies to $\epsilon v^2/c^2$, and therefore Eq. (3) informs us that

$$m_1-m_2=2\epsilon/c^2. \tag{4}$$

Therefore, notes Einstein, if a body gives off a quantity of energy in the form of radiation, then its mass diminishes by that energy divided by c^2. He then remarks with characteristic boldness that it makes no difference that the energy withdrawn from the body happens to have gone into radiation. This leads him to conclude that if the energy of a body changes for any reason, then its mass changes by the amount of the energy change divided by c^2.

Although his conclusion is independent of electrodynamics, Einstein had to invoke it at an intermediate step to determine how the energy of the radiation changed with a change of frame. We, being more at home with relativistic kinematics than Einstein's audience in 1905, can use virtually the same *Gedankenexperiment* to reach his conclusion without ever leaving the realm of mechanics. The knowledge of how the energy of radiation transforms is replaced by the velocity addition law, which is most conveniently expressed in the following form:

Let a body move uniformly with speed u. Then in a frame of reference moving with speed v at an angle θ to the direction of motion of the body, its speed u' satisfies[3]:

$$\frac{1}{\sqrt{1-u'^2/c^2}}=\frac{1-uv\cos\theta/c^2}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}. \tag{5}$$

Equipped with this addition law, we replace the two identical quantities of light in Einstein's argument by two identical particles, about which we know no more than we do about the particle originally under investigation. We thus consider a particle at rest with energy content E_1, that emits two identical particles, moving in opposite directions with the same speed u. After the emission the emitting particle remains at rest (as symmetry requires) with energy content E_2. The energy of each of the emitted particles is some function of its speed $E_3(u)$, the form of which is to be determined. Energy conservation requires that

$$E_1-E_2=2E_3(u). \tag{6}$$

Following Einstein, we next consider the same decay in a frame of reference that moves with speed v at an angle θ to the direction of emission of the particles. The initial and final energies of the emitting particle in that frame are $E_1(v)$ and $E_2(v)$. The emitted particles will have energies $E_3(u')$ and $E_3(u'')$, where u' and u'' are given by Eq. (5) with the two different choices of sign for $\cos\theta$. Energy conservation in the new frame requires that

$$E_1(v) - E_2(v) = E_3(u') + E_3(u'') . \qquad (7)$$

This relation must hold whatever the angle θ between the direction of motion of particles 3 and the new frame of reference. Since the left-hand side of Eq. (7) is independent of θ, we conclude that the quantity

$$E_3(u') + E_3(u'') \qquad (8)$$

cannot depend on θ.

Note that in Einstein's derivation this independence of θ is an immediate consequence of the transformation rules for the quantities of light and is explicit in Eq. (2). Indeed it is quite unnecessary to consider general θ to establish Einstein's argument (although he did); the case $\theta = 0$ by itself leads directly to the result. In our case, however, the requirement of θ independence imposes a strong constraint on the unknown function $E_3(u)$, which turns out to be enough to restore the ground lost by renouncing the use of electrodynamics. The argument is as follows:

Since the speed of particle 3 cannot exceed[4] c, we can equally well consider $E_3(u)$ to be a function of the variable $1/\sqrt{1-u^2/c^2}$:

$$E_3(u) = f(1/\sqrt{1-u^2/c^2}) . \qquad (9)$$

This change of variable is designed to take advantage of the fact that $1/\sqrt{1-u'^2/c^2}$ and $1/\sqrt{1-u''^2/c^2}$ depend linearly on $\cos\theta$ in the velocity addition law (5). Thus $E_3(u') + E_3(u'')$ becomes

$$f\left(\frac{1-uv\cos\theta/c^2}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}\right) + f\left(\frac{1+uv\cos\theta/c^2}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}\right) . \qquad (10)$$

This will evidently be independent of $\cos\theta$ if f is a linear function of its argument. More importantly, a little elementary analysis shows that if f is a continuous function of its argument (as an acceptable energy function must surely be) then Eq. (10) can be independent of $\cos\theta$ for frames moving with arbitrary speeds v *only* if f has the linear form $f(z) = az + b$.[5] This is our central mathematical result.

But if f has this form, then the energy of the particles 3 must depend on their velocity according to the law

$$E_3(u) = E_3 + k_3(1/\sqrt{1-u^2/c^2} - 1), \qquad (11)$$

where E_3 and k_3 are velocity-independent constants characteristic of particles of type 3.

The constant $E_3 = E_3(0)$ is the energy content of either of the particle of type 3 in its rest frame, and the constant k_3 determines the overall scale of its kinetic energy,

$$K_3(u) = E_3(u) - E_3 = k_3(1/\sqrt{1-u^2/c^2} - 1) . \qquad (12)$$

Since no special properties of particles 3 entered into the above argument, we may conclude that any type of particle must be characterized by constants E and k such that the velocity dependence of its energy is given by

$$E(u) = E + k(1/\sqrt{1-u^2/c^2} - 1) . \qquad (13)$$

In particular, using the form (13) for the emitting particle (with subscripts 1 and 2 for its pre- and post-emission states) we can write the equation of energy conservation in the moving frame [Eq. (7)] as

$$E_1 + k_1\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right) - E_2 - k_2\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right) = 2E_3 + 2k_3\left(\frac{1}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}} - 1\right) . \qquad (14)$$

By subtracting from Eq. (14) its form in the original ($v = 0$) frame, and canceling a factor of $1/\sqrt{1-v^2/c^2} - 1$ common to all terms, one extracts the following relation between the kinetic energy coefficients of the emitting particle before and after the emission:

$$k_2 = k_1 - 2k_3/\sqrt{1-u^2/c^2} = k_1 - 2k_3 - 2K_3(u) . \qquad (15)$$

This is our central physical result.

By examining the u dependence of the kinetic energy (12) when u is small compared with c we learn that in the nonrelativistic limit k/c^2 is called the mass m of the particle. If we adopt that traditional nomenclature, the content of (15) is that m_2 is less than $m_1 - 2m_3$ by $1/c^2$ times the kinetic energy with which the type 3 particles are expelled in the rest frame of the emitting particle. This is Einstein's conclusion, modified by the fact that in our version of the argument the emitted particles carry away mass as well as kinetic energy.

Thus if a particle in its rest frame decays into a collection of particles with a total kinetic energy K, then the total mass of the final group of particles must be less than the initial mass by K/c^2. Note that the mass that enters into this relation is the sum of kinetic energy coefficients k/c^2. If a particle gives up mass to produce kinetic energy K in such a process, then this loss is directly reflected in a lowering of the energy needed to accelerate the particle to a given velocity, as determined by its kinetic energy coefficient k. It is a remarkable fact that the simple relativistic kinematics of space and time measurements embodied in the addition law (5), imply by themselves that if energy is to be conserved at all, then it is necessary that the relation between the velocity and kinetic energy of a particle be subject to this kind of modification.

We emphasize the profound character of this conclusion, because the rest energy E in Eq. (13), which has not yet entered into the argument at all, is also conventionally assigned the numerical value mc^2. In contrast to the remarkable relation (15) obeyed by the kinetic energy coefficient, this other use of $E = mc^2$, though it is sometimes cited with comparable fanfare, has very little content. To see this we examine further the implications of energy conservation for the rest energies E.

Consider a group of particles, initially moving with velocities u_i with total energy

$$U = \sum_i E_i(u_i) = \sum_i \left(g_i + \frac{k_i}{\sqrt{1-u_i^2/c^2}}\right) , \qquad (16)$$

$$g_i = E_i - k_i . \qquad (17)$$

After some number of collisions, emissions, or mergers, the velocities, kinetic energy coefficients k_i, and the rest energies E_i may all have changed, but only in such a way that the total energy U remains unaltered. This conservation law must hold in all frames of reference. Since the k_i and E_i (and hence the g_i) are all invariant under a change of frame, only the velocities u_i depend on frame of reference.

The total energy U' in a frame moving with velocity v is therefore simply given by applying the addition law (5) to the velocities in Eq. (16):

$$U' = \sum_i \left(g_i + k_i \frac{1 - \mathbf{u}_i \cdot \mathbf{v}/c^2}{\sqrt{1 - u_i^2/c^2}\sqrt{1 - v^2/c^2}} \right). \qquad (18)$$

If U and U' are both conserved, then so is $U - U'\sqrt{1 - v^2/c^2}$. But this last quantity is just

$$\left(1 - \sqrt{1 - \frac{v^2}{c^2}}\right) \sum_i g_i + \mathbf{v} \cdot \sum_i \frac{(k_i/c^2)\mathbf{u}_i}{\sqrt{1 - u_i^2/c^2}}. \qquad (19)$$

The quantity (19) must be conserved whatever the velocity $\mathbf{v}$ of the moving frame. Since the first sum is even in $\mathbf{v}$ and the second odd, both must separately be conserved. The second sum is nothing but the relativistic momentum (with, note well, the kinetic energy coefficient of each particle divided by c^2 playing the role of its mass m); the first is just the sum of all the g_i.

We have therefore deduced (in the conventional way) that energy conservation implies momentum conservation, but also the conservation of something else. What are we to make of this? Note first that since $g = E - k$ is additively conserved, then because the additively conserved quantity $E(u)$ has the structure (13), the quantity $k/\sqrt{1 - u^2/c^2}$ is itself additively conserved, and as good a candidate for the energy function as the more general form (13). We can take advantage of this to redefine the energy function to be given by

$$E(u) = k/\sqrt{1 - u^2/c^2}, \qquad (20)$$

but we must then keep in mind that if there is an additively conserved Lorentz invariant associated with each particle, then E can be trivially redefined to include an additional velocity-independent term proportional to that quantity, without doing violence to the conservation laws.

This is as far as one can go without a more detailed dynamical theory. Electric charge, for example, is such a Lorentz invariant additively conserved quantity. So (except in grand unified theories) is baryon or lepton number. At the level on which we are operating there are no logical grounds *a priori* for excluding terms proportional to such quantities in the definition of rest energy. There is, however, something to be said for taking g to be zero in the definition of $E(u)$, resulting in the definition (20). For the choice $g = 0$ identifies the rest energy E of a particle with its kinetic energy coefficient k. The rest energy then becomes a measure of how much energetic capital a particle has available for the production of new kinetic energy. The conceptual price one pays for this convenience is the blurring of the distinction between the rest energy and the kinetic energy coefficient.

Note, finally, what happens when one tries to make the same argument directly from the nonrelativistic addition law $u'^2 = u^2 + v^2 - 2uv\cos\theta$. One immediately deduces the non relativistic version of Eq. (13): $E(u) = E + \frac{1}{2}mu^2$ (giving the kinetic energy coefficient its conventional name). The crucial result (15), however, is lost. One finds only $m_2 = m_1 - 2m_3$. Using this one deduces that $E_2 = E_1 - 2(E_3 + \frac{1}{2}m_3u^2)$, i.e., that the extra kinetic energy in the inelastic process must be subtracted from the rest energy of particle 2. In the nonrelativistic case this is nothing but a trivial bookkeeping device—the limit of the trivial relativistic convention about the rest energy. What is lacking in the nonrelativistic argument is any basis for making the nontrivial link between the creation of kinetic energy and the dynamically important kinetic energy coefficients, that gives $E = mc^2$ its deep physical content in the relativistic case.

APPENDIX A

Let the 4-velocity of the emitted particle be

$$u^{(4)} = (c,\mathbf{u})/\sqrt{1 - u^2/c^2}, \qquad (A1)$$

and the 4-velocity of the moving frame,

$$v^{(4)} = (c,\mathbf{v})/\sqrt{1 - v^2/c^2}. \qquad (A2)$$

The Lorentz invariant inner product of these is

$$u^{(4)} \cdot v^{(4)} = \frac{c^2 - \mathbf{u} \cdot \mathbf{v}}{\sqrt{1 - u^2/c^2}\sqrt{1 - v^2/c^2}}. \qquad (A3)$$

The value of this quantity is independent of frame. But in the moving frame v is 0 and u has, by definition, the value u', so that

$$u^{(4)} \cdot v^{(4)} = c^2/\sqrt{1 - u'^2/c^2}. \qquad (A4)$$

The addition law (5) is simply the statement that the evaluations (A3) and (A4) of the inner product are the same.

APPENDIX B

One can prove quite simply that the energy function f is linear by assuming that the dependence of energy on velocity is smooth enough to be twice differentiable. (In Appendix C we relax this to a simple assumption of continuity.) Since the energy (10) is independent of the angle θ, its second θ derivative (at fixed u and v) must vanish. This gives

$$f''\left(\frac{1 - uv\cos\theta/c^2}{\sqrt{1 - u^2/c^2}\sqrt{1 - v^2/c^2}}\right) + f''\left(\frac{1 + uv\cos\theta/c^2}{\sqrt{1 - u^2/c^2}\sqrt{1 - v^2/c^2}}\right) = 0. \qquad (B1)$$

Setting $\cos\theta = 0$ in (B1) and noting that the speed v of the moving frame can have any value between 0 and c, we learn that

$$f''(z) = 0, \quad z \geqslant 1/\sqrt{1 - u^2/c^2}. \qquad (B2)$$

Since u is fixed[6] Eq. (B2) by itself is not enough. To establish that f'' vanishes over the rest of its range set $\cos\theta = 1$ in Eq. (B1). With u and v positive, the second term in Eq. (B1) then vanishes as a consequence of Eq. (B2), and therefore so must the first:

$$f''\left(\frac{1 - uv/c^2}{\sqrt{1 - u^2/c^2}\sqrt{1 - v^2/c^2}}\right) = 0. \qquad (B3)$$

As v varies from 0 to u, the argument of f'' in (B3) goes from $1/\sqrt{1 - u^2/c^2}$ down to unity, giving us the vanishing of f'' in the rest of its range. It follows that throughout its entire range $1 \leqslant z < \infty$, f must be of the form $az + b$.

APPENDIX C

With a little more effort (and without the use of calculus), one can reach the conclusion of Appendix B, making only the weaker (and physically even more compelling) assumption that the energy function f is continuous. If the energy (10) is independent of $\cos\theta$ then it is equal to its value when $\cos\theta = 0$:

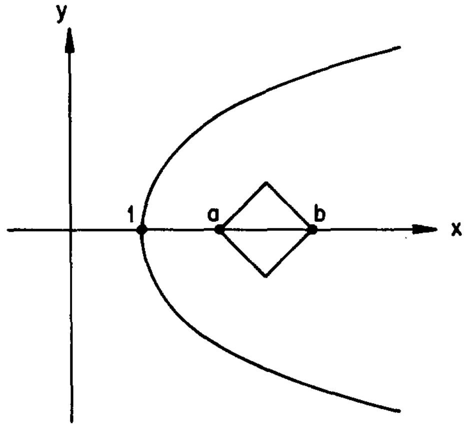

Fig. 1. The region specified in Eq. (C7) lies within the right-hand branch of the hyperbola $x^2-(c/u)^2y^2=1$. The square shown within that region contains all points (x,y) with $a\leqslant x\pm y\leqslant b$.

$$f\left(\frac{1-uv\cos\theta/c^2}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}\right)+f\left(\frac{1+uv\cos\theta/c^2}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}\right)$$
$$=2f\left(\frac{1}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}\right). \quad \text{(C1)}$$

We must show for fixed u, that if this condition holds for $0\leqslant v/c<1$ and $-1\leqslant\cos\theta\leqslant 1$, then $f(z)$ must be linear in z for all $z\geqslant 1$.

Note first that it is enough to establish that $f(z)$ is linear when $z\geqslant 1/\sqrt{1-u^2/c^2}$. For with $\cos\theta$ equal to unity, the arguments of the second and third occurrences of f in (C1) are both greater than or equal to $1/\sqrt{1-u^2/c^2}$; replacing f in both cases by $az+b$, one finds that

$$f(x)=ax+b \quad x=\frac{1-uv/c^2}{\sqrt{1-u^2/c^2}\sqrt{1-v^2/c^2}}. \quad \text{(C2)}$$

But as v goes from 0 to u, x goes from $1/\sqrt{1-u^2/c^2}$ to 1, thereby establishing that f has the form $az+b$ in the rest of its range.

Next define

$$f\left(\frac{t}{\sqrt{1-u^2/c^2}}\right)=g(t). \quad \text{(C3)}$$

We have

$$g\left(\frac{1-uv\cos\theta/c^2}{\sqrt{1-v^2/c^2}}\right)+g\left(\frac{1+uv\cos\theta/c^2}{\sqrt{1-v^2/c^2}}\right)$$
$$=2g\left(\frac{1}{\sqrt{1-v^2/c^2}}\right), \quad \text{(C4)}$$

and must show that $g(z)$ is linear for $z\geqslant 1$. With new variables

$$x=1/\sqrt{1-v^2/c^2},\quad y=(uv/c^2)\cos\theta/\sqrt{1-v^2/c^2}, \quad \text{(C5)}$$

Eq. (C4) becomes

$$g(x)=\tfrac{1}{2}[g(x+y)+g(x-y)] \quad \text{(C6)}$$

for all x and y satisfying

$$x\geqslant 1,\quad -(u/c)\sqrt{x^2-1}\leqslant y\leqslant (u/c)\sqrt{x^2-1}. \quad \text{(C7)}$$

The inequalities (C7) confine x and y to the interior of the right-hand branch of a hyperbola. Consider any square within this region, centered at a point on the x axis, and tilted at 45° so that diagonally opposite vertices of the square meet the axis at $x=a$ and $x=b$ (Fig. 1). If we can show for any such square that $g(x)$ is linear for $a\leqslant x\leqslant b$, then we shall have established that g is linear for all $x\geqslant 1$, since the entire x axis beyond $x=1$ can be covered by overlapping squares of this kind. Since the sides of the square are $x\pm y=a$, $x\pm y=b$, if (C6) holds for x,y in the square, then with $x+y=r$, $x-y=s$, we must also have

$$g[\tfrac{1}{2}(r+s)]=\tfrac{1}{2}[g(r)+g(s)], \quad \text{(C8)}$$

for all r and s is the interval $[a,b]$.

Equation (C8) asserts that the value of g midway between *any* two points in the interval $[a,b]$ is just the linear interpolation of the value at those two points. Consequently, if we have a set of points S from $[a,b]$ on which g is linear, then g will remain linear on S when we add to it the points midway between any two of the original points. Exploiting this fact, we can easily construct a dense set S of points in $[a,b]$ on which g is linear. We first put into S the two endpoints of the interval. We can then add to S the midpoint of the interval. Having done this we can add the remaining points that divide the interval into quarters, since they are halfway between the two ends and the midpoint. In the same way, we can next add to S the remaining points that divide the interval into eighths, then sixteenths, thirty-seconds, etc., all the while maintaining the linearity of g on S. In this way we will eventually add to S points arbitrarily close to any point in $[a,b]$. Therefore if g is continuous in $[a,b]$ it must be linear on the entire interval.

[1]A. Einstein, Ann. Phys. **18**, 639 (1905).

[2]For a critical discussion of many examples, see the references in P. C. Peters, Am. J. Phys. **54**, 804 (1986). Peters himself gives a discussion of the relativistic conservation laws that, like ours, derives and solves a functional equation for the form of the unknown conserved quantities. Peters finds the standard relativistic expressions by considering an elastic collision. Assuming that these forms continue to be conserved in inelastic collisions he then derives, along traditional lines, the dependence of mass on energy content.

[3]See Appendix A for a particularly simple derivation of Eq. (5).

[4]We appeal to causality; an appeal to dynamics would introduce a disquieting circularity into the argument. Note that with this change of variables Einstein's version of the *Gedankenexperiment* (with $u=c$) can only be viewed as a singular limiting case of ours.

[5]This is done in Appendix B under the stronger assumption that the energy function f is smooth enough to be twice differentiable. In Appendix C we give a slightly more elaborate argument (suggested by C. L. Henley) that requires only continuity (and has the further virtue of being accessible, if presented with care, to students innocent of the calculus.)

[6]In an earlier version of this argument, we allowed u to vary, but assumed that the particles 3 could be emitted in the same internal state whatever the value of the speed u at which they emerged. We are indebted to E. M. Purcell for pointing out the restrictive nature of this assumption, and questioning the need for it. The present argument requires only a single speed of emission, restoring the necessary flexibility by introducing a continuum of directions for the moving frame.