Energy and Momentum
Version A: Full Math

W. Blaine Dowler

April 21, 2012

Contents

1 What Is Energy?

One of the quantities that physicists find most useful is energy. It is often difficult to define for new students, due in large part to the fact that it is an intangible quantity that you cannot see, touch, taste or smell. The formal definition is often "energy is the ability to do work," which seems clear and straightforward until one is asked to define "work" in the scientific sense.

An alternative definition of energy is "the ability to cause changes in motion or position." From a scientific perspective, this is equivalent to the first definition, but it reduces the terminology to that which is intuitive for new students. There are different types of energy in the world, though we will only deal with two of them in this series.

1.1 Kinetic Energy

The type of energy that is often the easiest for students to accept is kinetic energy, or energy of motion. If an object is in motion, its own position is constantly changing. It also has the potential to cause changes in motion to other bodies through collisions.

In the Newtonian view, kinetic energy can be calculated based solely on the mass and speed of an object. If two objects with different mass travel at the same speed, then the object with greater mass has greater kinetic energy. If two objects with the same mass travel at different speeds, then the one with the greater speed has the greater kinetic energy. Moreover, the speed of the object has significantly greater impact than the mass. (If you double an object's mass without changing its speed, you double its kinetic energy. If you double its speed, you more than double its kinetic energy.)

1.2 Potential Energy

The second major type of energy is potential energy. This is a little harder to see, as there is nothing visibly moving or altered in any way as a result of obtaining potential energy. This energy can be "released" and turned into kinetic energy in one form or another. For example, a book on a shelf has gravitational potential energy: the book might fall, gaining speed and kinetic energy as it approaches the floor. A book on a lower shelf has less potential energy, as it cannot fall as far and will be traveling at a lower speed when it hits the ground.

Potential energy is a consideration when there is some sort of outside field or force that can drive changes in motion, such as gravity. As such, further discussion will be saved for the general theory of relativity. The variables one needs to describe potential energy depend upon the source and nature of the outside force or field.

2 What Is Momentum?

Momentum is a term that is commonly used in day to day speech in a fashion that differs from its scientific meaning. When people talk about momentum in the colloquial sense, they typically mean something is in motion and hard to stop. This usage is closer to the word "inertia" than momentum. It is inertia that determines the resistance to accelerations and changes in motion. Momentum, in the Newtonian view, is the product of mass and velocity.[1] In Newton's day, it was referred to as the "quantity of motion." Momentum is always *conserved* in a collision, meaning that the total momentum that goes into a collision is the same as the total momentum that comes after. (Again, direction matters; two vehicles meeting head on can turn into a stationary wreck immediately after collision because momentum in one direction, say North, can be canceled out by momentum in the opposite direction, South.)

3 How Are They Connected?

Kinetic energy and momentum can both be calculated using the same two pieces of information: the mass of the body in motion, and the velocity of the body in motion. Although they are not identical quantities,[2] the fact that they depend solely on the same two variables begs the question about whether or not the two are related.

3.1 The Newtonian View

In the Newtonian view, energy and momentum are connected through forces. When a force is applied to a object at rest, it accelerates to some velocity. The force applies over a period of time, during which the object travels some distance. Assuming the object is in space, so we don't have to worry about friction and

[1] It uses velocity, and not speed, because direction matters.

[2] They cannot be the same: direction matters to momentum but not kinetic energy, which is a significant difference on the conceptual side.

other factors, allowing the entire force to apply to accelerating the object in the exact direction of the force, then we can calculate both the energy and momentum using simple and similar math. We will also assume that the force is uniform, meaning it applies with the same strength and in the same direction the entire time it is applied. To calculate the energy of the object, one multiplies the force by the distance it has traveled. To calculate the momentum of the object, one multiplies the force by the time elapsed during the acceleration.[3]

3.2 The Relativistic View

In the relativistic view, there are a few changes that need to be made to these concepts. The first and foremost idea that needs to be explored is whether or not there is an upper limit to either quantity. After all, the universe has a speed limit in the speed of light. If these quantities depend on the speed of an object, then one would tend to think that there could be an upper limit to either or both in turn.

The relationship between the two quantities, as force applied along different directions (space or time), still applies in relativity. In that case, conceptually, we see that there should be no upper limit to either quantity: one can continually apply a force to an object, regardless of how much time elapses or distance is traveled. So, how do we obtain unlimited energy and momentum with a limited speed?

The answers lies in the fact that we have misidentified the variables involved. As we mentioned in lesson two, Newton originally recognized inertia as a key quantity, and then proposed that the mass and inertia of an object were one and the same quantity. Although he couldn't possibly have realized this at the time, mass and inertia are not, in fact, the same. The inertia of an object increases as its speed increases. This increase is almost imperceptible at the speeds that we see in our everyday experiences, but it is unmistakeable at higher speeds. As an object gets closer to the speed of light, its inertia increases incredibly rapidly, ensuring that no amount of energy or momentum will carry the body above the speed of light. In several of the cases in which we used mass in Newtonian mechanics, we should have used inertia as Newton originally proposed. This concept was so counter intuitive, and the notion of equality between mass and inertia so widely accepted, that when this difficulty was first noticed physicists incorrectly assumed that the mass of an object was increasing, and the quantities were named as such. The mass was referred to as *rest mass* and the inertia was referred to as *relativistic mass*. Although the error in these terms is now understood, they were in use for so long that they still appear in much of the literature on the subject.

[3] If the force is not uniform, some adjustments need to be made. Instead of simple multiplication, we use calculus to integrate the force with respect to the other variable.

The next question pertains primarily to momentum: with momentum, direction matters, so what quantity appears as the time component of momentum? We now have four directions that need to be described when dealing with quantities that use direction at all. To discover this component, we begin with Newton's definition of momentum as the product of inertia and velocity. We take our four dimensional velocity and multiply it by the inertia of an object. What we find in the time component of the result is startling: it is the kinetic energy of the object, with an additional term added: mc^2, where m is the mass of the object and c is the speed of light.

This was utterly shocking. An object at rest contained energy in the absence of outside fields or forces, and that energy E had the quantity mc^2. Never before had it occurred to anyone that mass could somehow store energy, and yet it seemed apparent that mass was some form of condensed, solidified energy. This revelation is the one that led to nuclear power. It was so unexpected that Einstein himself discussed it tangentially in his paper, deviating from the calculation he worked on to make a special point of describing it. It is also one of the primary examples of serendipitous discovery in the sciences: who would have believed that nuclear energy would have been discovered by examining momentum in a universe with a speed limit?

4 Newtonian Calculations

To better prepare the contrast between Newtonian calculations and relativistic calculations, we shall review the Newtonian formulae of mechanics and show how they are derived.

4.1 Force

The expression for force is one of the few *axiomatic* formulae in classical mechanics. As an axiomatic formula, it has no derivation; it is simply believed to be true. Newton proposed force as the product of inertia and acceleration. He soon proposed that an object's mass was identical to its inertia, resulting in the formula

$$\mathbf{F} = m\mathbf{a} = m\frac{d\mathbf{v}}{dt} = m\frac{d^2\mathbf{x}}{dt^2}$$

Much of the rest of mechanics are drawn from these formula.

4.2 Momentum

Momentum was proposed through the relationship

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$

In other words, force causes an object to change momentum over time. In cases in which the mass of the accelerated object remains constant, this reduces simply though an integral to

$$\mathbf{p} = \int d\mathbf{p} = \int \mathbf{F} dt = \int m \frac{d\mathbf{v}}{dt} dt = \int m d\mathbf{v} = m\mathbf{v}$$

which is the version people are most accustomed to seeing. In cases with variable mass, such as any self-propelled, fuel burning object, the final expression becomes more complicated, but it is still based on integrating the applied force over time.

4.3 Energy

Energy E is defined scientifically as the ability to do work W, or to change the state of motion of an object. There are two types of energy, primarily defined through the concepts of force and work. The two forms are kinetic and potential energy, and kinetic energy is the easiest to derive. Both are based on the idea that

$$W = \Delta E = \int \mathbf{F} \cdot d\mathbf{x}$$

In other words, applying a force in the direction of motion of an object changes its state of motion and does work, changing its energy.

4.3.1 Kinetic Energy

The kinetic energy derivation is the simplest to do in many ways. Again, the familiar form is based on the idea that the mass (or inertia) of an object remains constant during the motion of the body. If an object starts with zero energy and is accelerated to velocity $\mathbf{v}$, then the kinetic energy E_k is given by

$$E_k = \int \mathbf{F} \cdot d\mathbf{x} = \int m \frac{d\mathbf{v}}{dt} \cdot d\mathbf{x} = \int m d\mathbf{v} \cdot \frac{d\mathbf{x}}{dt} = \int m d\mathbf{v} \cdot \mathbf{v} = \frac{1}{2} m v^2$$

which is the familiar form.

4.3.2 Potential Energy

Potential energy is a bit more difficult to recognize. Imagine you ride an elevator up three floors in a building. We know energy is being applied, even through the middle portion of the ride which is at a constant speed, as the elevator car continues to climb and consume electrical power. The net change in speed, however, is zero: you were motionless before the elevator started to climb, and you were motionless when you reached the top. This is a sign that your potential energy has changed: were you to step into the elevator shaft when no elevator was present, gravity would quickly cause you to accelerate to ground level. Similar situations can be developed when charged particles are close to each other, or magnets are close to each other, and so forth. Any time one might experience a force of some kind, there is a potential involved.

One of the most common potentials we face is gravity. In the case of first exposure to this force, we tend to work with problems at or near the surface of the planet Earth, where the acceleration due to gravity is given by $g = 9.81\text{m/s}^2$ in the downward direction.[4] This value of g remains fairly constant near the planet's surface. Because of this near constancy, we can calculate the potential energy of an object a height h from the planet's surface by examining the action of the force of gravity as the object falls.

$$\Delta E_p = \int \mathbf{F} \cdot d\mathbf{x} = \int m\mathbf{g} \cdot d\mathbf{x} = \int_h^{0\text{m}} mgdz = -mgh$$

This is almost the familiar form. The negative sign appears because this calculation represents the change in potential energy as an object drops *from* a height h to ground level; the object has *lost* mgh worth of potential energy. By convention, ground level is taken to be the point at which potential energy is zero, so this means the object had a positive mgh worth of energy when it was still at height h. Thus,

$$E_p = mgh$$

is the familiar expression for an object at height h above the Earth's surface. This works for the balls dropped off bridges and buildings in most introductory problems, but doesn't work very well for objects in orbit. With distances that great, g cannot be treated as a constant. In those cases, we work with the full gravitational force expression

$$F_g = \frac{Gm_1m_2}{r^2}$$

where m_1 and m_2 are the masses of the two objects involved, $G = 6.67 \times 10^{-11}\text{Nm}^2/\text{s}^2$ is Newton's gravitational constant, and r is the distance between

[4]In fact, when one tries to come up with a scientific definition of the word "down," one is forced to define it as the direction that the force of gravity points.

the centres of mass of the two objects. Integrating over r results in the expression

$$E_p = \int \mathbf{F}_g \cdot d\mathbf{r} = \int \frac{Gm_1m_2}{r^2}dr = -\frac{Gm_1m_2}{r}$$

The negative sign appears because of a convention that results from something known as *gauge freedom*. When doing these experiments, we often decide arbitrarily where the point of zero height is in our apparatus. When dropping things in a lab, it is convenient to choose the floor, the lab bench, or whatever other target we are using as the point of zero energy. We do this no matter which floor of the building we are on, knowing full well that the object we drop could continue to build up kinetic energy if we were to drop it down the stairs or into a floor drain instead. We have the freedom to set the "zero" of our gauge to whatever level is most convenient; the integral definition we use emphasizes the fact that it is the *change in* position that is important, and not the position itself. So, where is our zero energy point when dealing with situations too big to fit in a lab room, such as an orbiting satellite? We set the zero energy point at infinity, making potential energies *negative* as one object accelerates and approaches the other object.[5] This also means that an object that starts at rest at infinity and falls towards the other body arrives with kinetic and potential energies that still add up to zero, so that $E_k = -E_p$, which makes other calculations convenient.

Of course, further expressions could be developed for other forces. For example, a complete electromagnetic potential could be developed by integrating the Lorentz force law:

$$E_p = \int q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) d\mathbf{x}$$

but these are not directly related to what is coming next.

5 Relativistic Calculations

5.1 Time Derivatives and Invariants

In order to work effectively in relativity, we need to ensure that our four vectors are Lorentz invariant. We have established for our position four-vector already, but we still need to confirm this for our four dimensional velocity and its timelike component.

[5] Technically, both accelerate towards each other, but in most real life problems, we typically treat the planet or star as an object at a fixed point in space and the satellite as moving. The motion of the planet or star is generally too small to be measured or relevant.

Let us begin with the instinctive option:

$$\vec{u}' = \frac{d\vec{x}'}{dt'} = \frac{d}{dt'}\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \frac{d}{dt'}\begin{pmatrix} \gamma\left(ct - \frac{v}{c}x\right) \\ \gamma\left(x - vt\right) \\ u_y \\ u_z \end{pmatrix} = \begin{pmatrix} c \\ \frac{u_x - v}{1 - \frac{u_x v}{c^2}} \\ u_y \\ u_z \end{pmatrix}$$

Comparing this to

$$\vec{u} = \frac{d\vec{x}}{dt} = \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix}$$

we see that this is only invariant when

$$u_x = \frac{u_x - v}{1 - \frac{u_x v}{c^2}}$$

which can be manipulated to the point

$$v u_x^2 = v c^2$$

This is true in only two situations. If $v = 0$, this is true, but hardly useful; that means S and S' are the same reference frame. If $v \neq 0$, then we have $u_x = \pm c$, which is hardly useful in general situations. Thus, our instinct for an invariant four-vector fails us, and we need to use another option.

The problem lies in the definition's right hand side $\frac{d\vec{x}'}{dt'}$. The numerator is the differential of an invariant quantity, while the denominator is not. To correct this, we need to differentiate with respect to another invariant quantity. Thankfully, this was introduced in the previous lesson, though it did not seem that significant at the time.

With most time derivatives, one must differentiate with respect to the proper time, and not the observer's time, when performing calculations. In essence, we are treating the S' frame as the reference frame in which the moving body is at rest. In short, we are setting $\vec{v} = \vec{u}$ and using this as our reference. With this definition,

$$c^2\tau^2 = -c^2t^2 + x^2 + y^2 + z^2$$

is an invariant quantity. When the object is at rest, we find

$$d\tau = dt\sqrt{1 - \frac{u^2}{c^2}} = \frac{dt}{\gamma\left(u\right)}$$

which is, essentially, our time dilation equation for $u = v$.

This leads us to the four vector definition

$$\vec{u} = \frac{d\vec{x}}{d\tau} == \frac{d\vec{x}}{dt}\frac{dt}{d\tau} = \gamma\left(u\right)\begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix}$$

Now we calculate $\vec{u} \cdot \vec{u}$ to test invariance, noting that $v = u$:

$$\begin{aligned} \vec{u} \cdot \vec{u} &= \gamma \begin{pmatrix} c & u_x & u_y & u_z \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \gamma \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix} \\ &= \gamma^2\left(-c^2 + u_x^2 + u_y^2 + u_z^2\right) \\ &= \frac{-c^2 + u^2}{1 - \frac{u^2}{c^2}} \\ &= -c^2\frac{1 - \frac{u^2}{c^2}}{1 - \frac{u^2}{c^2}} \\ &= -c^2 \end{aligned}$$

The final result for this quantity is independent of any information about $\vec{u}$ in any way. Thus, invariance results.

When acceleration is involved, things are sadly more complicated. Invariance is true when dealing with an inertial reference frame. We cannot use the S' frame as the rest frame for an accelerated object, as this is *not* an inertial reference frame. The very nature of acceleration shows this to be the case, making any hope of invariance rather slim. Thus, we define the acceleration as

$$\vec{a} = \frac{d\vec{u}}{d\tau} = \frac{d^2\vec{x}}{d\tau^2}$$

with little hope of an invariant quantity. In fact,

$$\vec{a} = \frac{d\vec{u}}{d\tau} = \frac{d\vec{u}}{dt}\frac{dt}{d\tau} = \gamma\frac{d\vec{u}}{dt} = \gamma^2 \begin{pmatrix} \frac{ua\gamma}{c} \\ u_x\frac{ua\gamma^2}{c^2} + a_x \\ u_y\frac{ua\gamma^2}{c^2} + a_y \\ u_z\frac{ua\gamma^2}{c^2} + a_z \end{pmatrix}$$

If we look specifically at the rest frame of the object in the exact moment the acceleration is first applied, we find that $\mathbf{u} = \mathbf{0}$ and $\mathbf{a} = \alpha$, which we can refer to as the proper acceleration, or the acceleration at measured by the object at rest. In this frame,

$$\vec{a} = \begin{pmatrix} 0 \\ \alpha_x \\ \alpha_y \\ \alpha_z \end{pmatrix}$$

As it turns out, this leads to an invariant quantity:

$$\vec{a} \cdot \vec{a} = \vec{a'} \cdot \vec{a'} = \alpha^2$$

Thus, even though it seemed unlikely that we'd reach an invariant, we have, further justifying our use of $d\tau$ as the preferred time differential for relativity.

5.2 Force

The relativistic force is relatively easy to compute. As with Newtonian mechanics, it is the product of mass and acceleration. We define the four-vector force as the four-vector acceleration multiplied by the mass of the object:

$$\vec{F} = m_0 \vec{a} = m_0 \frac{d\vec{u}}{d\tau}$$

The subscript on the mass m_0 is there because we will soon find that mass and inertia are not identical as was believed in Newton's day. This form also assumes that mass is constant, which is not necessarily the case. The invariance of this form is relatively simple to establish: $\vec{F} \cdot \vec{F} = m_0^2 \vec{a} \cdot \vec{a} = m_0^2 \alpha^2$, which is invariant.

In the case in which the mass is variable, the new form becomes

$$\vec{F} = \frac{d\left(m_0 \vec{u}\right)}{d\tau}$$

5.3 Momentum

Now that we have

$$\vec{F} = \frac{d\left(m_0 \vec{u}\right)}{d\tau}$$

for our force, we can build our momentum in an analogous fashion to the way we did in the Newtonian view. In that case,

$$\vec{F} = \frac{d\left(m_0 \vec{u}\right)}{d\tau} = \frac{d\vec{p}}{d\tau}$$

Thus,

$$\vec{p} = \int \vec{F} d\tau = \int \frac{d\left(m_0 \vec{u}\right)}{d\tau} d\tau = \int d\left(m_0 \vec{u}\right) = m_0 \vec{u}$$

which probably would have been our instinctive choice to begin with.

Now, doing this with algebra is all well and good, but what do the components mean? The spatial components of $\vec{p}$ are easy enough to interpret, but now we have a time component as well. What is the natural analogue to this component?

Let us examine this four-vector explicitly.

$$\vec{p} = \begin{pmatrix} \gamma m_0 c \\ \gamma m_0 u_x \\ \gamma m_0 u_y \\ \gamma m_0 u_z \end{pmatrix}$$

When the object is at rest and $u = 0\text{m/s}$, we are left with

$$\vec{p} = \begin{pmatrix} m_0 c \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

We see that the time component is the mass of the object multiplied by the speed of light, but what does that *mean*? This is some sort of "momentum through time," but what is a momentum through time? The quantity must be important in some way, as we find that the invariant length of our momentum vector is closely related:

$$\vec{p} \cdot \vec{p} = -m_0^2 c^2$$

So, this combination of mass and the speed of light must be important, but we need to see why that is.

5.4 Energy

In the three dimensional world, we knew that

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$

and

$$F = \frac{dE}{dx}$$

If the same relationships hold in relativistic mechanics, then we would have something analogous to

$$\frac{dp}{dt} = \frac{dE}{dx}$$

to relate the two quantities. Dealing with this in four dimensional space, we have

$$E = \int d\vec{p} \cdot \frac{d\vec{x}}{d\tau} = \int \vec{u} \cdot d\vec{p}$$

In the case when $\mathbf{u} = 0\text{m/s}$, we are left only with the time components of these vectors:

$$E = m_0 c^2$$

which may be the single most famous equation in the history of science. Furthermore, this is a natural occurrence of the time component of momentum. Our "momentum through time" is nothing more than the energy of the object divided by the speed of light. Even more surprisingly, objects at rest with no potential energy still have energy by virtue of mass! Performing the calculation in the general form gives us that

$$E = \gamma m_0 c^2 = mc^2$$

where

$$m = \gamma m_0$$

is the inertia of the object in motion. Thus, energy and inertia are revealed to be closely related aspects of the same quantity. It was this discovery, coupled with quantum mechanics, that led to the development of nuclear technology.

5.4.1 Kinetic Energy

If the energy of an object at rest is

$$E = m_0 c^2$$

and the total energy of an object in motion (with no potential energy) is given by

$$E = \gamma m_0 c^2$$

then it stands to reason that the kinetic energy of an object is the difference between these two:

$$E_k = \gamma m_0 c^2 - m_0 c^2 = m_0 c^2 \left(\frac{1}{\sqrt{1 - \left(\frac{u}{c}\right)^2}} - 1 \right)$$

This looks nothing at all like the familiar

$$E_k = \frac{1}{2}mv^2$$

of the Newtonian world that works so well in the high school lab. How can it possibly be correct?

Those who have taken calculus may be familiar with Taylor series expansions. The basic concept is that any continuous function with continuous derivatives can be approximated by a (possibly infinite) polynomial. If one takes the relativistic kinetic equation above and applies this technique, one finds that the infinite approximating polynomial begins

$$E_k = \frac{1}{2}m_0u^2 + \frac{3}{8}m_0\frac{u^4}{c^2} + \dots$$

The first time is the formula we are used to. The next term is one that is divided by c^2, which is no small number. In fact, each term is of the form

$$\frac{p}{q}m_0\frac{u^{k+2}}{c^k}$$

where k, p and q are all whole numbers with $p < q$. As a result, when $u \ll c$, the relativistic nature of the inertia is difficult to detect, and the Newtonian form is an excellent approximation.

It is also worth noting that the relativistic kinetic energy formula derived here is one that allows for unlimited energy, but contains an implicit speed limit.

5.4.2 Potential Energy

When objects have a significant amount of potential energy, they tend to experience a variety of accelerations. Most potential energy formulae are virtually unchanged (aside from the replacement of mass with inertia). These will be developed more fully in lessons seven through nine.