The Special Theory of Relativity is based on two postulates:

1. The laws of physics have the same form in all inertial systems.

2. The speed of light in a vacuum is the same for all observers.

Mathematically, this means that the laws of physics must be formulated in terms of physical quantities which obey the *Lorentz transformations*. The basic building blocks to construct such quantities are *four-vectors*. A four-vector $V^\mu = (V^0, V^1, V^2, V^3) = (V^0, \vec{V})$ has 4 components: one "time-like" component V^0, and three "space-like" components (V^1, V^2, V^3) or $\vec{V}$ which are analogous to ordinary vectors in Newtonian physics. The Lorentz transformations then state that if one observer measures the quantities V^μ, a second observer who moves at constant velocity v in a particular direction (say, the x-direction) will measure the quantities V'^μ, given by

$$V'^0 = \gamma \left(V^0 - \frac{v}{c} V^1 \right) \tag{1}$$

$$V'^1 = \gamma \left(V^1 - \frac{v}{c} V^0 \right) \tag{2}$$

$$V'^2 = V^2 \tag{3}$$

$$V'^3 = V^3 \tag{4}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{5}$$

is the Lorentz factor. Also, the product

$$V = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} V^\mu V^\nu \tag{6}$$

defines a quantity, called a *Lorentz scalar* which remains invariant under Lorentz transformations; indeed you can verify that $\sum_{\mu,\nu} \eta_{\mu\nu} V^\mu V^\nu = \sum_{\mu,\nu} \eta_{\mu\nu} V'^\mu V'^\nu$. Here,

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{7}$$

is called the *Minkowski metric*. The most fundamental four-vector is the space-time position $x^\mu = (x^0, x^1, x^2, x^3) = (ct, \vec{x}) = (ct, x, y, z)$. With the infinitesimal displacements $\mathrm{d}x^\mu$ we can then define the Lorentz scalar $\mathrm{d}\tau$, called the *proper time*, as

$$c^2\mathrm{d}\tau^2 = -\sum_{\mu,\nu}\eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \tag{8}$$

$$= \left(c^2 - v^2\right)\mathrm{d}t^2. \tag{9}$$

Therefore, $\mathrm{d}t = \gamma\mathrm{d}\tau$. The proper time allows us to construct the *velocity four-vector* u^μ

$$u^\mu = (u^0, \vec{u}) = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \left(c\frac{\mathrm{d}t}{\mathrm{d}\tau}, \frac{\mathrm{d}\vec{x}}{\mathrm{d}\tau}\right) = (\gamma c, \gamma\vec{v}), \tag{10}$$

and the *momentum four-vector* p^μ

$$p^\mu = (p^0, \vec{p}) = mu^\mu = (\gamma mc, \gamma m\vec{v}). \tag{11}$$

The spatial part of p^μ is obviously a relativistic version of the Newtonian momentum. But what is the temporal component γmc? To answer this, consider a system on which no external forces act. Analogous to Newtonian physics, the total four-momentum of this system will be constant. Again the spatial components correspond to conservation of ordinary momentum. But we know from classical physics that the *energy* of the system will also be conserved. Therefore, the additional temporal constant is nothing else than the conservation of energy, up to a factor c. Thus

$$E = cp^0 = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}}. \tag{12}$$

As a side note, the conservation of momentum and energy are closely related to symmetries: if a system behaves the same regardless of where it is (translation invariance), momentum is conserved; if it behaves the same regardless of when we observe it (time invariance), energy is conserved. This further justifies the fact that p^0 is indeed proportional to E.

If we set the velocity $\vec{v} = 0$, the energy reduces to $E_0 = mc^2$. So, unlike the Newtonian case, a system that isn't moving still possesses an internal energy, related to its mass. Also, from the Lorentz scalar

$$\sum_{\mu,\nu}\eta_{\mu\nu}p^\mu p^\nu = -E^2/c^2 + |\vec{p}|^2 = -\gamma^2 m^2 c^2 + \gamma^2 m^2 v^2 = -m^2c^2, \tag{13}$$

we obtain the relation

$$E^2 = m^2c^4 + |\vec{p}|^2c^2. \tag{14}$$

For small velocities, the energy reduces to

$$E = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \approx mc^2 + \frac{1}{2}mv^2 + \dots. \tag{15}$$

We see that the second term is the non-relativistic kinetic energy. Finally, you can verify that

$$\mathrm{d}\vec{p} \cdot \vec{v} = \mathrm{d}(\gamma m \vec{v}) \cdot \vec{v} = \left(\gamma^3 m \frac{v^2}{c^2} + \gamma m\right) \vec{v} \cdot \mathrm{d}\vec{v} = \gamma^3 m \vec{v} \cdot \mathrm{d}\vec{v} = \mathrm{d}E, \tag{16}$$

which is the relativistic version of the work-energy theorem in classical physics.