Getting its from bits

Frank Wilczek

The inventor of the term 'black hole', John Wheeler, has a gift for memorable phrases. 'Getting its from bits' is another of his creations. It refers not to an object, but to a vision of a world derived from pure logic and mathematics. That vision has to a remarkable extent been embodied in modern physics — here is a progress report.

The 'its from bits' programme[1] has a venerable history, for perhaps the first great quantitative generalization in science was Pythagoras' discovery of the numerical patterns behind musical sounds. When two strings of a lyre — of the same material, and under equal tension — are played together, they produce a pleasant harmony precisely when their lengths are a ratio of small integers: 2 to 1 for an octave, 3 to 2 for a musical fifth, 4 to 3 for a fourth, and so on. For the followers of Pythagoras, this provided a satisfying example of a principle they held to be completely general, the idea that 'all is number'.

A chain of thought extending over two millennia links this idea to the inspirations of Kepler. Kepler's three laws of planetary motion are enshrined in textbooks, and provided the foundation for Newton's celestial mechanics. Less publicized is his erroneous 'zeroth' law, which was his version of Copernicanism, and the point of departure for his original research. According to Kepler's zeroth law, which would have pleased Pythagoras, the orbits of the six planets are great circles on spheres alternately inscribed within and circumscribed about the five regular solids. Of course, we now know that there are more than six planets, and Kepler himself was reluctantly forced, by Tycho Brahe's accurate observations, to abandon circular orbits in favour of ellipses. According to modern views, the number of planets and the size of planetary orbits was determined more or less accidentally during the complicated process whereby our Solar System condensed out of a gigantic interstellar gas cloud. Solar systems around other stars, which are now beginning to yield their secrets to observation, are expected to be very different.

Indeed, classical physics teaches us that the size of planetary orbits is not the sort of thing we should aspire to predict. It makes a sharp distinction between the basic laws, which govern the evolution of systems in time, and are expected to be simple, and the initial conditions, which must be given from outside. The equations of classical physics can be applied to any number of different types of solar system, having different sizes and shapes. There is nothing in Newton's laws of gravity and mechanics, nor for that matter in the other pillar of classical physics, Maxwell's electrodynamics, that could serve to fix a definite size. Symptomatic of this, there is no way to form a characteristic length from the parameters that govern these theories, namely the gravitational coupling G and the speed of light c. Classical physics is profoundly anti-Pythagorean.

Box 1: A few words on dimensional analysis

Dimensional analysis is a time-honoured way to estimate the answer to a physical question without having to solve or even, perhaps, to fully formulate the governing equations. The main idea is both trivial and profound. It is that physical results must be independent of the choice of units.

A classic application of dimensional analysis is to fluid flow. Suppose we are interested in flow of velocity v past a body of size L, in a fluid of viscosity per unit density ν. Since ν has dimensions of length^2/time, while v of course has dimensions of length/time, the only dimensionless quantity we can form is the Reynolds number, $\mathrm{Re} = vL/\nu$. So, for example, we can use model aircraft in a wind tunnel to study the flow around real aircraft, by compensating with a larger v for a smaller L. As long as Re stays the same, the flows will differ only by trivial re-scalings.

In its abstract form, dimensional analysis leads to the vague, but extremely useful, principle that reasonably defined quantities should be numbers of the order unity when expressed in appropriate 'natural units'. In the system of natural units used in particle physics, quantities having dimensions of mass, length and time are given the dimensions of powers of energy (usually in electronvolts), which effectively makes Planck's constant of action $\hbar$ and the speed of light, c, both equal to unity. Essentially all the equations in particle physics contain the speed of light c and Planck's constant of action $\hbar$, so it is natural to regard these as fundamental units. Planck proposed that in a complete formulation of physics, not yet attained, the only additional parameter to appear would be Newton's gravitational constant G. In such a theory, all other constants of nature, expressed in these Planck units, would be calculable pure numbers.

F. W.

Modern Pythagorism

The most successful theory of modern physics, quantum mechanics, completely changes the situation. Quantum mechanics provides a unique ground-state configuration for each atom and molecule, thus relieving the indeterminacy in the analogous classical theory of solar systems, and making it possible to understand why atoms and molecules exhibit well-defined, universal chemistry. At a yet deeper level, quantum field theory, which is the logical extension of quantum mechanics to include special relativity, explains why the elementary constituents — electrons and nuclei — exist in myriads of identical copies, each being an excitation of a single universal field. So, for example, quantum electrodynamics (QED) posits, in addition to the familiar electromagnetic field whose excitations represent the formation of photons, an electron field whose excitations we see as the creation of electrons.

When Planck introduced his quantum of action, $\hbar$, he immediately advertised the possibility of a new Pythagorism[2]. The inability of classical physics to provide a definite scale of length had been relieved. For Planck observed that from G, c and $\hbar$ one can form the Planck length:

$$l_{\mathrm{Planck}} = \left[\frac{G\hbar}{c^3}\right]^{1/2} \sim 10^{-33}\,\mathrm{cm}$$

More generally, by combining appropriate powers of these parameters one can reproduce any unit of measurement needed in the description of the physical world. On the other hand, one cannot combine them to produce a dimensionless pure number. Thus G, c and $\hbar$ provide an ideal, non-redundant system of physical units. From this arises the modern Pythagoras–Planck programme: to formulate a theoretical framework in which G, c and $\hbar$ are all profoundly incorporated, and to calculate within that framework all the constants of nature, expressed in Planck's units, as pure numbers.

This is a tall order. A particular challenge is that fundamental quantities such as the size of atoms or the mass of the proton turn out to have outlandish values (about 10^{25} and 10^{-19}, respectively) when expressed in Planck units. If the Pythagoras–Planck programme is to succeed, then the standard working assumption of dimensional analysis (see Box 1), that naturally defined entities should be of order unity in natural units, must be profoundly subverted.

As atomic physics developed, some of the spirit of the Pythagoras–Planck programme was realized, but major compromises were required. For many purposes it is a very good approximation to neglect the effects of relativity, and to regard nuclei as infinitely heavy compared with electrons. In this approximation, the fundamental equations of atomic and molecular physics can be formulated in a way that $\hbar$, together with m_e and e, the mass and charge of the electron, appear as the only parameters. From these we can construct a unique unit of length, the Bohr radius:

$$a_r = \frac{\hbar^2}{e^2 m_e}$$

This does give the approximate size of atoms — so in this case dimensional analysis is vindicated.

In a more accurate treatment of atoms and molecules (such as QED) one must include relativistic effects, and the ability of finite mass protons and nuclei to recoil. The description of these effects brings c, and the finite masses of the proton and other atomic nuclei, into the equations. (Gravity is utterly negligible here, so G is not required.) Once c is added to the parameters of atomic physics, one can form the fine-structure constant, α, a dimensionless quantity:

$$\alpha = \frac{e^2}{\hbar c} \sim 0.00735$$

This parameterizes the strength of the electromagnetic attraction between protons and electrons, or equivalently the size of the quantum of electric charge. In the spirit of Planck and Pythagoras, one should not be satisfied to have such a quantity appearing as fundamental in the laws of physics. Rather, one should aspire to calculate it.

The pioneers of atomic physics were acutely aware of this challenge. Pauli was fond of saying that the first question he would ask the Almighty would be to explain the value of the fine structure constant. (The joke continues, that after hearing the explanation — from Satan — Pauli thought for a moment, then snapped "Wrong!".) The challenge escalates when we consider the nuclear masses. Indeed, by taking ratios of these masses, or the ratio of any of them to the electron mass, we can construct many more dimensionless numbers. To satisfy Pythagoras and Planck, we would have to calculate all these numbers, not just take them from experiment.

Strong insights

The modern theory of the strong interaction, which binds atomic nuclei together, is quantum chromodynamics (QCD). This theory has notably advanced us towards getting 'its from bits', in three distinct ways. First, it accounts, in principle, for those problematical nuclear masses. A full formulation of QCD requires, on the face of it, seven parameters: a pure number α_s, analogous to the fine-structure constant, that governs the strength of the strong interaction, plus the masses of six different types of quarks, in addition to $\hbar$ and c. The up, down, strange, charm, bottom and top quarks are the particles that, together with the colour gluons, carry the colour charges of QCD. Although even this would be reasonably economical, considering the amount of data to be correlated, that parameter count is grossly unfair to QCD. The up and down quark masses are very small, and they are the only two quarks that are significant for nuclear physics. By putting their masses to zero, and ignoring the other quarks, one obtains an excellent approximate theory containing just one dimensionless quantity, α_s.

In practice it is very difficult to use QCD to calculate nuclear masses, just as it is very difficult to do self-contained calculations of chemical processes beginning with the Schrödinger equation of quantum mechanics. We have faith in the theory primarily because it is able to give an accurate and detailed account of high-energy processes, where the calculations become much simpler (see Fig. 1). There has also been impressive progress in calculating the masses and properties of the mesons and baryons that take part in strong interactions. These are analogous to atoms formed of quarks, antiquarks and gluons, whereas nuclei — aside from the proton itself — are analogous to complicated molecules. Representative results are shown in Fig. 2 (for a fuller discussion, see Box 2 and ref. 3). These results leave little doubt that correct values of the nuclear masses would emerge from more numerical work, but definitive calculations are probably some years off.

Second, QCD brings to the fore a profound property of quantum field theories, what we might call the relativity of charge. According to modern quantum physics the vacuum, which evolution has selected us to regard as an empty background, is in reality a highly structured, responsive and dynamic medium. Because of the uncertainty principle the 'vacuum' contains virtual particles that can, like the molecules in an insulator, arrange themselves to partially screen an inserted charge. If that happens, the charge one measures at smaller distances, inside the screening cloud, or equivalently in higher-energy processes, will effectively increase. The opposite behaviour, antiscreening or

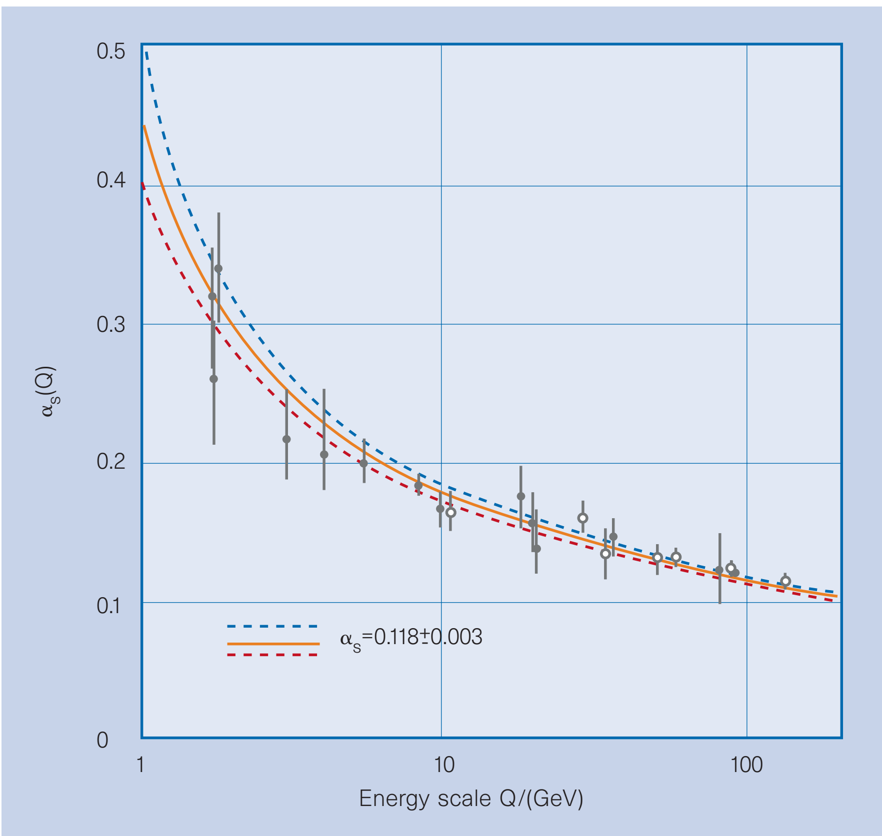

Figure 1 **The relativity of charge. Value of the strong coupling constant, α_s, established by a variety of experiments (data points) at different energy scales, and compared to the QCD theoretical prediction for α_s (solid line). See ref. 4 for detailed references to the experiments.**

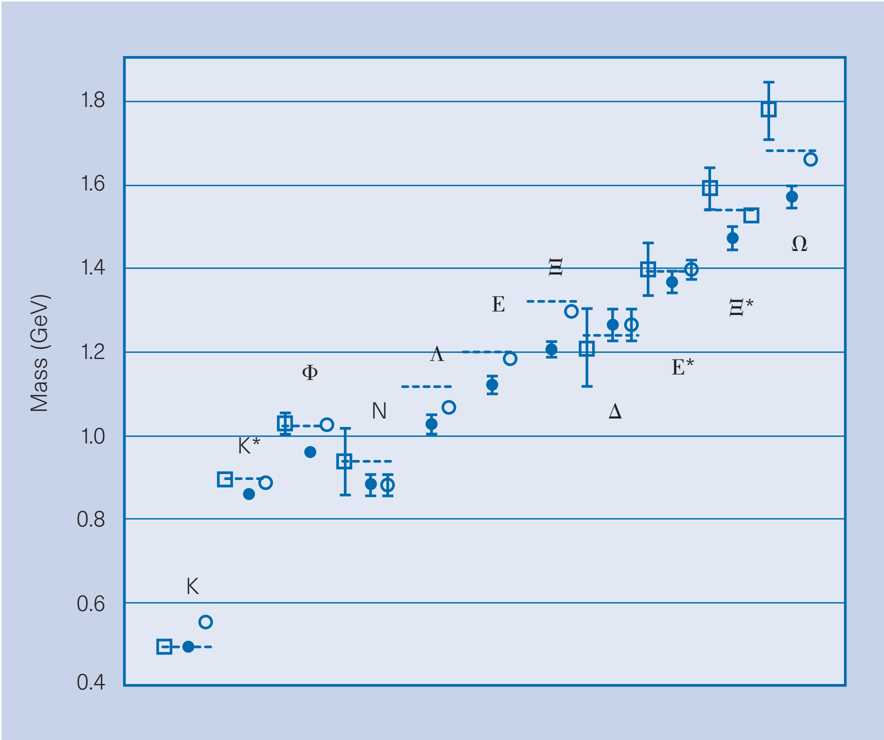

Figure 2 **Comparison of masses of light hadrons (dotted lines) to various lattice simulations (data points). These calculations contain just one free parameter, the strange quark mass. Sources of error in the current lattice calculations, which are believed to be responsible for the small residual errors, are discussed in ref. 3.**

asymptotic freedom, though less familiar, is also possible. In either case, the value of the charge, or coupling strength, is not an absolute concept, but depends on how it is measured. Antiscreening is calculated to occur in QCD. The experimental evidence for this behaviour is now quite firm[4], as you can see in Fig. 1.

Because of the relativity of charge, the QCD analogue of Pauli's question — why is the value of the fine structure constant what it is? — receives a startling answer: "It's anything you like, at some distance or other". We can simply declare it to be, say, 1/10, thereby defining the distance where it is 1/10. This is the phenomenon of dimensional transmutation[5]. A dimensionless measure of the quantum of charge, the coupling 'constant', has been transmuted into a unit of distance.

The approximate QCD theory with two massless quarks appears, naively, to be a family of theories, each with a different value of the coupling, and none defining a scale of distance. But because of dimensional transmutation, it turns out to be a family of perfectly identical theories that differ only in the units they use to measure length. This difference in units matters for comparison of purely QCD quantities to non-QCD quantities, such as the ratio of the diameter of the proton to the Bohr radius, but it does not affect dimensionless quantities within QCD itself, such as ratios of nuclear sizes or nuclear masses.

So QCD, in its slightly idealized version with two massless quarks, provides a truly marvellous partial realization of the vision of Pythagoras and Planck. Using $\hbar$ and c as units, and with no further inputs — except the number of colour charges, of which there are three (binary '11'), and the number of quarks, of which there are two (binary '10') — it accurately accounts for all the 'its' of nuclear physics, and much else besides. 'Its from bits', to be sure!

Getting it all — or hitting a wall?

Although QCD accounts admirably for the strongest forces in nature, it is certainly not a Theory of Everything. What, if anything, does it portend for the full Pythagoras–Planck programme? This brings me to my third and final point. The relativity of charge, which plays such a central role in QCD, applies as well to the other interactions of the Standard Model of modern physics — the weak and electromagnetic interactions (although for them it is a much smaller effect). This brings up the possibility that all the couplings — that is the quanta of each of the strong, weak and electromagnetic charges — might have a common value when measured at exceedingly small distance scales (or equivalently at high energies), despite their disparate values at currently accessible scales. There are several other pieces of evidence pointing toward this possibility, as I described in these pages last year[6].

For our present discussion, what is crucial is that the inverse couplings depend logarithmically on the distance at which they are measured. So, a small coupling will evolve only very slowly. As an illustration, the strong coupling α_s is observed to change from a value close to 1 at 10^{-13} cm to about 1/8 at 10^{-16} cm, and is predicted to be about 1/25 at 10^{-33} cm. So, whereas the strong coupling might eventually merge with its weaker brethren, its approach is quite a drawn-out affair. When we calculate where the unification takes place, we find a truly remarkable result. The strong, electromagnetic and weak couplings, which are significantly different when measured at 'practical' distances, are calculated to become equal when measured at distances about 17 orders of magnitude smaller — near the Planck unit of distance.

It is extraordinarily suggestive that the Planck scale emerges here. To appreciate why, we must consider extending the notion of the relativity of charge to gravity. The sorts of charges, strong, weak or electromagnetic, to which the interactions of the Standard Model of particle physics respond, change only logarithmically with distance, owing to subtle quantum mechanical effects. But gravity responds to energy directly, so that it runs linearly with energy (or inverse distance) scale. From its much inferior strength at accessible energies, gravity ascends to equality with the other interactions at roughly the Planck scale. Thus we discover that all the coupling strengths become equal simultaneously. Even in the absence of a detailed theory, we find here a concrete, semi-quantitative indication that all of the basic forces arise from a common source.

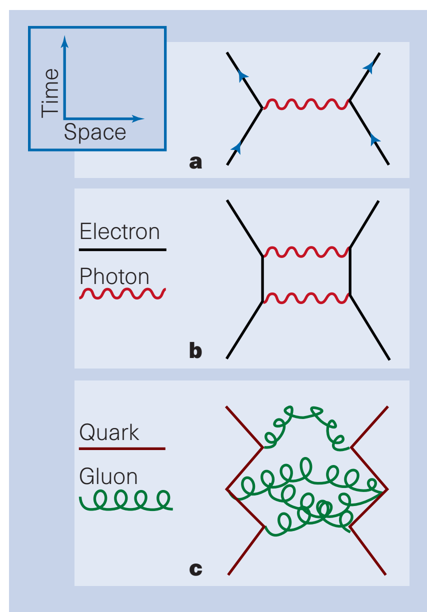

Figure 3 **Feynman graphs.** a, **The simplest graph contributing to electron–electron scattering in QED, by exchange of a virtual photon. In a more accurate calculation,** b, **one must allow for multiple exchanges.** c, **A typical contribution to the interaction of quarks in QCD.**

This example of how vastly different scales emerge, provides critical insight for the vexing problem, fundamental for the Pythagoras–Planck programme, of how to generate extremely large (or extremely small) dimensionless numbers. Any dynamical effect due to a large coupling automatically generates an exponentially large ratio of scales, if the fundamental coupling (before antiscreening) is small. An outstanding example is the proton mass, or equivalently its quantum size (Compton wavelength) defined by $\hbar / m_{\mathrm{proton}}c$. According to QCD, the proton's Compton wavelength is essentially determined by the dimensionally transmuted length where the strong interaction becomes strong and holds in the quarks. That occurs when α_s is measured to be unity. But this is related, by the relativity of charge, to the exponentially smaller distance where unification takes place. Putting this idea into an equation, we find:

$$m_{\mathrm{proton}} \sim \exp(-k/\alpha_{\mathrm{unified}}) M_{\mathrm{Planck}}$$

for the proton mass in Planck units. Here, $M_{\mathrm{Planck}} = (\hbar c/G)^{1/2} \approx 10^{18}\, m_{\mathrm{proton}}$ is the Planck mass unit, $\alpha_{\mathrm{unified}} \approx 1/25$ is the common value of the strong, electromagnetic and weak couplings when they unify, and $k = 11/2\pi$ is a calculable numerical factor that characterizes the antiscreening. This formula works remarkably well. Suddenly one sees 'outlandish' numbers like 10^{25} from the perspective of $\exp(-1/\alpha)$ — which is actually considerably bigger — and they no longer appear quite so daunting.

Although all of these developments justify optimism, it remains conceivable that the 'its from bits' programme will hit a wall. A particularly serious possibility is that we will converge on a unique set of basic equations for physics — many physicists believe that such equations will emerge from investigations into superstring theory — but that these equations will contain consistent solutions describing many basically different possible worlds. There might, for example, be valid solutions describing worlds with different electron/proton mass ratios, or different numbers of quarks. Twenty years ago, one might have objected, against this possibility, that if there were other solutions, we should have seen regions of the known Universe where they are realized. After all, the Universe is a very big place (volume ~10^{180} in Planck units). But inflationary cosmology[7], which posits that the entire observable Universe expanded from a small patch early on, has made it plausible that the known Universe is homogeneous not for any fundamental reason, but just because we are only sampling a small patch of reality. With this in mind, we need only travel sufficiently far, or wait sufficiently long, to encounter differences of the sorts mentioned above. Many particulars of what we commonly regard as the most basic features of the world would then hinge on an accident of history (that is, which amplified patch we emerged from). Attempts to calculate the electron mass from first principles might be as futile as attempts to calculate the shape of the Solar System, or the anatomy of frogs. Still, we must try. □

Frank Wilczek is at the Institute for Advanced Study, School of Natural Sciences, Olden Lane, Princeton, New Jersey 08450, USA.
e-mail: wilczek@sns.ias.edu

1. Misner, C., Thorne, K. & Wheeler, J. in *Gravitation* Ch. 44 (Freeman, New York, 1973).
2. Planck, M. *S.-B. Pruss. Akad. Wiss.* 440–480 (1899).
3. Burkhalter, R. http://xxx.lanl.gov/abs/hep-lat/9810043
4. Schmelling, M. http://xxx.lanl.gov/abs/hep-ex/9701002
5. Coleman, S. & Weinberg, E. *Phys. Rev. D* **7,** 1888–1910 (1973).
6. Wilczek, F. *Nature* **394,** 13–15 (1998).
7. Linde, A. *Inflation and Quantum Cosmology* (Academic, San Diego, 1990).

Box 2: Bit proliferation – crunching the numbers

If simple input parameters are to give extremely complex outputs, there must be a lot of logical processing inbetween. Here I describe the computational machinery that processes '11' and '01' into tables of nuclear properties.

The traditional way to visualize and calculate in quantum field theories is by means of Feynman graphs, which follow the tracks of particles in space and time. Particles that are not observed – those that are neither in the initial nor final state – are virtual particles. Feynman graphs describe the scattering of particles, by considering all the possible ways they can interact by exchanging virtual particles. Figures 3a and 3b show the first simple Feynman graphs for electron–positron scattering in quantum electrodynamics (QED). It is almost always a good approximation in QED to use only the simplest possible graph to describe the interaction, and an excellent approximation to use only a few. In quantum chromodynamics (QCD), on the other hand, the probability that more complicated graphs such as Fig. 3c contribute to the interaction is not particularly small. When many complicated graphs make substantial contributions, the sums become impractical.

An entirely different approach is necessary. The particle picture, epitomized by Feynman graphs, is an easy-to-calculate approximation for limited purposes, but the fundamental equations of QCD are formulated in terms of fields filling space and time. Indeed the simplest, and perhaps the most profound, way to state the theory is to give the rule which governs the probability amplitudes for different configurations of the fields. This rule is easily stated mathematically, is very symmetrical, and relates only the fields at nearby space–time points (that is, it is local). The difficulty is that when one applies the rule, one finds that many different configurations occur with substantial probability. They reflect that there are always violent, but short-ranged and short-lived, quantum fluctuations in the colour version of electric and magnetic fields, even in what evolution has designed us to regard as 'empty' space (for otherwise we'd always be distracted).

So far, no one has found a painless way to add these all up. The only really successful approach has been to crunch the numbers (see ref. 3 for a review). To do that, one first replaces continuous space–time by a lattice, and restricts attention to a finite box. The details are very intricate and clever, but one must check that the approximations involved in discretizing and boxing are not too severe. In practice, about 10^5 points are used, to ensure accuracy at the few per cent level. Sums over so many variables cannot be done analytically, so Monte Carlo sampling techniques are employed. Heroes working on numerical QCD have pushed the frontier of high-speed parallel processing, often designing and constructing their own computing machines. At the moment, two different teraflop machines are devoted full time to QCD calculations (1 teraflop = 10^{12} floating point multiplications per second).

Within this framework, one calculates the masses of observed hadrons, mesons, baryons or, in principle, nuclei and 'glueballs' (bound states made purely of colour gluons) by dropping appropriate mixtures of quarks, antiquarks and gluons into the roiling medium of fields at one space–time point, and measuring how long they hang together, and how fast they move. The particles we see are the resonant modes, which can persist as coherent entities for a reasonable amount of time. In these calculations, the masses of hadrons are found, quite literally, as the frequencies one can sound on an exotic gong, constructed to purely mathematical specifications. It is a result that would surely have pleased Pythagoras. **F.W.**