J. Phys. A: Gen. Phys., Vol. 5, August 1972. Printed in Great Britain. © 1972.

Time variation of the coupling constants

P C W DAVIES
Institute of Theoretical Astronomy, University of Cambridge, Cambridge CB3 0EZ, UK

MS received 25 January 1972, in final revised form 27 April 1972

Abstract. The possibility of the variation with time of the electromagnetic and strong coupling constants is investigated using nuclear systematics. Previous work by Dyson, and Broulik and Trefil, is re-examined and extended, with the conclusion that Teller's hypothesis of a logarithmic time dependence of the fine structure constant is apparently within the limits discussed if there is a corresponding logarithmic time dependence of the strong coupling constant also. Moreover the recent discovery of naturally occurring ^{244}Pu places the Gamow hypothesis of $e^2 \sim t$ much nearer the allowable limits than had previously been supposed.

1. Time variation of the coupling constants

Several authors have suggested that various fundamental quantities in physics that are normally regarded as constants are in fact variable over cosmological time scales. On the other hand it is occasionally remarked that many qualitative aspects of the universe are delicately dependent upon the values of these quantities (Dyson 1971). Consequently, we may use this delicate balancing to place rather strong limits on the variation rate of some quantities.

In this paper we restrict ourselves to discussion of the strong and electromagnetic coupling constants, g_s and e respectively. Variation in the latter quantity has been proposed by Gamow (1967), to explain the cosmological redshift:

$$e^2 \sim t$$

and Teller (1948):

$$\frac{1}{e^2} \sim \ln(mt).$$

Two conspicuous examples of systems whose properties are sensitive to the values of g_s and e are the atomic nucleus, and stellar interiors undergoing nucleosynthesis. Limits on the variation of e or g_s using nuclear systematics have been discussed by Dyson (1967), Peres (1967) and Broulik and Trefil (1971), and observational limits on e have been given by Bachall and Schmidt (1967). In this paper the nuclear argument will be examined in more detail and improved limits will be obtained for mutual variations in g_s and e. The weak coupling will be assumed constant throughout. Units such that $\hbar = c = 1$ will be used unless stated to the contrary, so that $e_0^2 \simeq 1/137$, $g_s^2 \simeq 15$ (zero subscripts or superscripts always denote the present measured values. The dimensionless ratios e/e_0 and $g_s/g_s^{(0)}$ will be denoted by E and G respectively.)

2. Nuclear stability

2.1. The two nucleon system

The stability of a given nuclide depends on a balance between three contributions to the energy—the attractive nuclear force, the nucleon degeneracy pressure (zero point energy) and the disruptive Coulomb repulsion. In light nuclei the Coulomb part is negligible, but dominates in heavy nuclei.

As an illustration consider the simplest system of all—the two nucleon problem. Each nucleon of reduced mass $\frac{1}{2}M_N$ can be considered to move in some sort of short ranged potential of depth V and width b. The zero point kinetic energy is given by the uncertainty principle as $\pi^2/4M_Nb^2$. If V falls off faster than $1/b^2$ it is clear that the system need not be bound. Suppose we consider a square well of constant radius b_0. Then the two nucleon state will be bound if

$$Vb_0^2 > \frac{\pi^2}{4M_N} \simeq 5{\cdot}2 \times 10^{-14}\,\text{cm}. \tag{1}$$

We have to decide on the values V and b_0. The deuteron can exist in a triplet S state, for which the values are (1 MeV $= 5{\cdot}1 \times 10^{10}\,\text{cm}^{-1}$ in our units)

$$V = 38{\cdot}5\,\text{MeV} = 2{\cdot}0 \times 10^{-11}\,\text{cm}^{-1}$$

$$b_0 = 1{\cdot}93 \times 10^{-13}\,\text{cm}$$

so that $Vb_0^2 \simeq 7{\cdot}3 \times 10^{-14}$ cm and the deuteron is just bound (experiment gives for the binding energy a value of 2·2 MeV $= 1{\cdot}1 \times 10^{11}\,\text{cm}^{-1}$).

In contrast, for the singlet state

$$V = 13{\cdot}3\,\text{MeV} = 6{\cdot}8 \times 10^{11}\,\text{cm}^{-1}$$

$$b_0 = 2{\cdot}58\,\text{fm} = 2{\cdot}58 \times 10^{-13}\,\text{cm}$$

giving $Vb_0^2 \simeq 4{\cdot}5 \times 10^{-14}$ cm so that the dineutron and diproton, which can only exist in the singlet state, are just unbound (experiment gives for this a negative binding energy of -92 keV $= -4{\cdot}7 \times 10^9\,\text{cm}^{-1}$). In the latter case there is an additional small Coulomb energy of $+0{\cdot}66$ MeV $= 3{\cdot}4 \times 10^{10}\,\text{cm}^{-1}$ (Okamoto and Pask 1971).

It is clear that with the total energy $H \simeq 0$ the existence or nonexistence of a bound two nucleon system is critically dependent on the value of g_s. A change in the nuclear force of a few per cent either way would be sufficient to unbind the deuteron, seriously affecting the hydrogen burning properties of stars, or to bind the diproton. Dyson (1971) has pointed out the importance that the diproton be unbound, the reason being that it is unstable against decay to form a deuteron: ${}^2\text{He} \to {}^2\text{H} + e^+ + \nu$. Now the main sequence hydrogen burning process of stars is $p + p \to {}^2\text{H} + e^+ + \nu$ and this proceeds about 10^{18} times slower than the deuteron burning process ${}^2\text{H} + {}^2\text{H} \to {}^3\text{H} + n$. Thus if the diproton were bound, all the hydrogen would have been burned up catastrophically at the beginning of the universe.

To estimate the dependence of H on g_s we can use the Yukawa theory as a first approximation. This describes a potential energy at internuclear separation r given by $g^2r^{-1}\exp(-M_\pi r)$. The effective coupling constant g is related to g_s by

$$g \simeq \frac{M_\pi}{M_N} g_s \simeq \frac{1}{7} g_s \tag{2}$$

where M_π and M_N are the pion and nucleon masses respectively, so that $g_0^2 \simeq 0{\cdot}3$. The separation distance r can be Bohr quantized to give, for the ground state:

$$\text{kinetic energy} = \frac{1}{M_N r^2} = \frac{1}{2} g^2 \exp(-M_\pi r)\left(\frac{1}{r} + M_\pi\right). \tag{3}$$

The total interaction energy H is then given by

$$H = \frac{1}{2} g^2 \exp(-M_\pi r)\left(M_\pi - \frac{1}{r}\right). \tag{4}$$

If we let $M_\pi \to 0$ and substitute from (3) into (4) we obtain the usual hydrogen atom binding energy formula.

From our considerations using the square well potential we know $H \simeq 0$ so that from (4) we have $1/r \simeq M_\pi$. In this region we may neglect the variation in the exponential factor and put $1/r = M_\pi + \epsilon$. Rewriting (3) and (4) we have

$$\frac{5{\cdot}44 M_N^{-1}}{r^2} = g^2\left(\frac{2}{r} + \epsilon\right) \tag{5}$$

$$H = -0{\cdot}18\epsilon g^2. \tag{6}$$

Substituting for ϵ from (5) into (6)

$$H = \frac{1}{r}\left(\frac{M_N^{-1}}{r} - 0{\cdot}37 g^2\right). \tag{7}$$

Using the fact that $H \simeq 0$ and $r \simeq M_\pi^{-1}$ at the present epoch, equation (7) gives $g_0^2 \simeq M_\pi/0{\cdot}37 M_N = 0{\cdot}4$ which is well within the accuracy to which we are working. Replacing g^2 with $Gg_0 = 0{\cdot}4G$ in (7) we may solve for the values of G which give zero binding energy, using the known present values of the range r. For the ^{3}S state we take 1·18 fm and for the ^{1}S state 0·97 fm. Using these values of r in (7) we have two equations for G

$$\begin{aligned} G^2 &= 1 + 8\times 10^{-13} H_t \\ G^2 &= 1 + 7\times 10^{-13} H_s \end{aligned} \tag{8}$$

where $-H_t$ and $-H_s$ are the currently observed values of the triplet and singlet state binding energies respectively. For the ^{3}S state we see from (8) that $G^2 = 0{\cdot}91$, that is, a decrease in G of about 5% is sufficient to unbind the deuteron. For the ^{1}S state $G^2 = 1{\cdot}003$, but in the case of the diproton we must add a Coulomb term to the second equation (8)

$$G^2 = 1{\cdot}003 + 0{\cdot}031\, E^2. \tag{9}$$

At the current value of $E = 1$, this means that a mere 2% increase in G is sufficient to bind the diproton.

2.2. *The many nucleon system*

When we come to consider the heavier nuclides we cannot be sure how the binding energy will depend on g_s or even if g_s is a significant parameter, in the absence of a good theory of nuclear forces. Such a theory would be very complicated. Each nucleon moves in some sort of potential well due to the average interaction of all the others.

Both the radius and depth of this well will depend on g_s, in some uncertain way, so rather than take as a model some experimentally determined well shape, we must base our considerations on some sort of dynamical model such as the Yukawa theory. For the heavier nuclides, the ratio of kinetic to nuclear potential energy is in the region of $\frac{1}{2}$ rather than 1. We regard each nucleon as moving in an average potential which is not too different from the Coulomb form. (For small internucleon separations we may neglect the exponential factor in the Yukawa potential.) We may then treat the system in simple analogy to the hydrogen atom and expect the binding energy to vary like g_s^4. Note that this is in contrast with the unjustified assumption of a g_s^2 dependence by Broulik and Trefil. This moderately strong nuclear binding is now competing against a strong Coulomb force. With increasing atomic weight, the individual nucleons become progressively more weakly bound as the Coulomb force dominates. For heavy nuclei, fragmentation becomes energetically possible. Estimates of the stability limits may be made using the well known semi-empirical mass formula, which includes contributions from the important competing effects. It gives an expression for the binding energy B†

$$B(AZ) = 15{\cdot}7A - 17{\cdot}8A^{2/3} - \frac{0{\cdot}71Z^2}{A^{1/3}} - 94{\cdot}8\frac{(Z-\frac{1}{2}A)^2}{A}. \tag{10}$$

The integers A and Z are the total number of nucleons and protons, respectively. The various coefficients are estimated from experiments.

Disintegration is energetically possible if $\Delta B > 0$, where ΔB is the difference in binding energy between parent and daughter nuclei. It follows from (10) that

$$\Delta B = -4{\cdot}6A^{2/3} + 0{\cdot}26Z^2A^{-1/3} \tag{11}$$

for the case of two equal mass fragments. If we define a fissionability parameter $\lambda = Z^2/A$ it follows from (11) that disintegration is possible for $\lambda \gtrsim 18$. However, this is incomplete, because in order for the particles to separate to infinity they must contend with a potential barrier caused by the change of Coulomb energy as the nucleus becomes distorted. This barrier may be surmounted for $\lambda \gtrsim 44$, a result which follows from the liquid drop model of the nucleus by considering instabilities against deformation, in the case of instantaneous fission. There will still be fission after a finite lifetime for any nuclei satisfying $\lambda \gtrsim 18$. But in practice, only nuclei with $\lambda \gtrsim 35$ are spontaneously fissile with observable lifetimes. If one considers nonsymmetric disintegration such as alpha decay, then the theoretical limit is $\lambda \simeq 30$, and the observable limit is $\lambda \simeq 33$. A variation of g_s enters the mass formula (10) via the coefficients in the first and second terms. This will change both the range of nuclei which are fissile, and their lifetimes. For alpha decay, the lifetime is exceedingly sensitive to changes in the energy of the emitted particle. The complexities of the nuclear many body problem are great, and we cannot be sure how a decrease in g_s will change the lifetimes, even for symmetric fission. However, it is possible to make some rough estimates of the basic stability of nuclei.

In figure 1 we have plotted Z against G for three values of λ (E constant) assuming the dependence $Z^2/A = G^4$. The curves obtained divide the stable and unstable regions for nuclei. The lower curve assumes $\lambda = 30$ and only corresponds to the situation where a decrease in g_s drastically reduces the lifetimes for decay. It is chosen to intersect the Z axis at $Z = 75$, which is the theoretical limit for the possibility of alpha decay

† We have used the more familiar MeV units here as only ratios will be required.

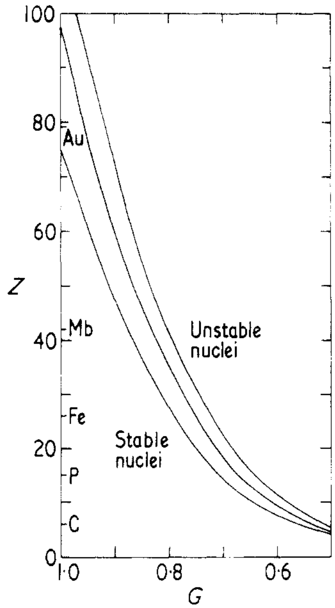

Figure 1. Nuclear stability as a function of the strong coupling (Coulomb forces constant).

(although symmetric fission may take place far below this value). The middle curve has $\lambda = 38$ and intersects at $Z = 98$, or around the region where fission is observed with halflives $\simeq 10^9$ yr. The upper curve is the extreme case of instantaneous fission with $\lambda = 44$. The most realistic curve must lie below this, and is probably in the region of the middle one.

Various elements have been marked for reference. It is seen that for a decrease in g_s of only about 25% biologically important elements like iron become unstable. Less than a 50% decrease would be expected to seriously affect the stability of even carbon, although the validity of the above mentioned models for such light nuclei is in some doubt.

2.3. *β decay*

If we regard a heavy nucleus as a box of degenerate gas then the neutrons and protons will fill up the energy levels to their respective Fermi energies E_n and E_p. When we switch on the Coulomb forces, E_p is raised to $\simeq E_n$, so that for all but the lightest nuclei $n > p$. If E_n is only slightly greater than E_p it is not energetically favourable to replace a neutron by a proton because, although the overall Fermi energy is decreased, there is an additional Coulomb energy from the extra proton which more than compensates. In this region it may be favourable for a proton to change into a neutron by positron emission or K capture. For E_n sufficiently greater than E_p however, the system will undergo β decay by electron emission. If the energy difference between the two nuclides is Δ, then (for $\Delta > 0$), this can be shown to occur with a lifetime which increases at least as fast as $\Delta^{-\alpha}$ where $\alpha = [2+\{1-(Z/137)^2\}^{1/2}]$ at $\Delta = 0$. Any variation of E or G will affect Δ and hence the lifetime. To estimate the dependence of Δ on E and G we consider the nucleus as a spherical box of radius r_0. The extra Coulomb energy is obtained simply by averaging the extra proton over the nucleus, and gives a factor $\propto Z^2E^2/r_0$. The loss in Fermi energy follows from a simple consideration of level densities, and gives a factor $\propto 1/r_0^2\{(Z/A)^{2/3}-(N/A)^{2/3}\}$. By our approximation of assuming a Bohr type theory for small nucleon separation, we may say $r_0 \sim g_s^{-2}$. We then have

$$\Delta = A_1G^4-A_2G^2E^2 \tag{12}$$

where A_1 and A_2 are chosen such that when $E = G = 1$, Δ is the observed binding energy difference. They may of course be estimated from the formula above. In considering mutual variations of E and G in the region of unity, we may approximate Δ to zero and write $E \propto G$ for no change in Δ.

3. Constraints on time variation

We may now use these results to discuss estimates of constraints on a possible temporal variation of E and G.

Broulik and Trefil (1971) have noted that there are several long ($\simeq 10^7$ yr) halflife transuranic elements that are not found naturally on earth. They argue that if E were less (or G greater) in the past, there may have been a time when these elements were stable against α decay. But we know that this time cannot have been less than about ten times their halflife ago, or they would be found in detectable abundance now. The authors choose ^{244}Pu with an observed halflife of about 7×10^7 yr for the best candidate, and arrive at the constraint (e constant)

$$\frac{1}{g_s^2}\frac{\mathrm{d}g_s^2}{\mathrm{d}t} \lesssim 2\times10^{-11}\ \mathrm{yr}^{-1}.$$

This element was an unfortunate choice, for four months later it was reported that ^{244}Pu had in fact been found to occur naturally (Hoffmann *et al* 1971). The explanation for this is thought to be unusual chemical concentration of the element. The lesson here is that it is always dangerous to argue about the nonexistence of something. Of course we could argue that the discovery is an indication in favour of varying coupling constants, but, as we shall see, greater values of G and lesser values of E of this magnitude can be ruled out by other examples.

A second criticism of the Broulik–Trefil work is that to estimate their constraint they assume that ^{244}Pu would be stable if its value for λ were changed to that for ^{238}U, which has a halflife in excess of 10^9 yr. The trouble with this is that it assumes the value of λ is more than a rough guide to stability. In fact, many lighter nuclei than ^{238}U have much shorter lifetimes. Indeed, it is the most stable nucleide above ^{209}Bi. This indicates that the unusual stability of ^{238}U is due to shell closure effects, and makes it a particularly bad candidate for λ value arguments. In the author's opinion we must assume that only when ^{244}Pu has a λ value in the region of the heaviest truly stable nuclei can we safely use this argument. This procedure would lead to the value $(1/g_s^2)\,\mathrm{d}g_s^2/\mathrm{d}t \lesssim 13\times10^{-11}\ \mathrm{yr}^{-1}$, but as we have mentioned, this is assuming a g_s^2 dependence of the binding energy. With the more rigorous g_s^4 dependence we obtain, finally, $(1/g_s^2)\,\mathrm{d}g_s^2/\mathrm{d}t \lesssim 6\times10^{-11}\ \mathrm{yr}^{-1}$.

We may invert this type of argument to put a limit on possible lower values of G in the past by noting that a sufficient decrease would tend to increase the instability against fission (either symmetrically or by α decay). This argument is subject to the same cautions as the ^{244}Pu case, and we take as a rough guide the same limit, $(1/g_s^2)\,dg_s^2/dt \gtrsim -6\times10^{-11}\,yr^{-1}$.

A greatly improved estimate of the former limit comes from the diproton argument, which (for constant E) requires that g_s^2 be within about 4% of its present values at the beginning of the universe ($\simeq 10^{10}$ yr ago). Then

$$\frac{1}{g_s^2}\frac{dg_s^2}{dt} \lesssim 4\times10^{-12}\,yr^{-1}.$$

The strongest constraints are those on E (for fixed G) given by Dyson (1967), who discusses the β active isotope ^{187}Re and its stable daughter isotope ^{187}Os. ^{187}Re has an observed halflife of 4×10^{10} yr with $\Delta = 2{\cdot}6$ keV, and $\alpha = 2{\cdot}835$. If in the past Δ were somewhat greater, or appreciably negative, then either all the ^{187}Re would have decayed by electron emission, or the ^{187}Os by K capture, and would not be observed in their present abundance in nature. Dyson gives the limits

$$\frac{1}{e^2}\frac{de^2}{dt} \lesssim 3\times10^{-13}\,yr^{-1} \qquad (\text{no } ^{187}\text{Re})$$

and

$$\gtrsim -11\times10^{-13}\,yr^{-1} \qquad (\text{no } ^{187}\text{Os}).$$

Finally, Bachall and Schmidt have observed the fine structure in the spectra of distant radio galaxies at a redshift of 0·2. Such galaxies are usually interpreted to be at about 2×10^9 light years, but at this distance no fine structure variation was found, to about 2 parts in 10^3, giving a limit on $(1/e^2)\,de^2/dt$ of about $10^{-12}\,yr^{-1}$. This observation does not depend on the value of g_s and appears to rule out the Gamow cosmology decisively (which requires $(1/e^2)\,de^2/dt \simeq 10^{-10}\,yr^{-1}$), but not quite the Teller cosmology, which only requires $(1/e^2)\,de^2/dt \simeq 10^{-12}\,yr^{-1}$. It should, however, be pointed out that lately all redshift interpretations are open to some doubt.

In all the other examples we should really consider mutual variation of G and E, and so the results are presented graphically in figure 2. In interpreting this graph one should be most cautious in view of the approximations used. It can serve as a rough guide only. Figure 2 shows lg G plotted against lg E in the region of $E = G = 1$. The box in the middle has a scale magnification of 10. The cross at the centre denotes the present value of E and G. In the Gamow cosmology, the cross would stand to the left of this position in the past, being somewhat beyond the left hand edge of the graph at the time of formation of the earth. Its position in the Teller cosmology would be slightly to the right at this time, and is denoted by a point marked T. We may consider in either case that the cross may also wander vertically through values of G as it moves horizontally.

In the central box, the shaded area between the two lines is the allowable region constrained by the β decay arguments. For constant G, we have used Dyson's limits. In the coarser scale, this elongated region appears as a thick line given from equation (12) as

$$\Delta = 0{\cdot}0026 = (15{\cdot}8+0{\cdot}0026)G^4 - 15{\cdot}8G^2E^2$$

where we have chosen the numbers (in MeV) such that $G = 1$ when $E = 1$, and have used Dyson's estimate of 15·8 MeV for the Coulomb energy difference. Neglecting

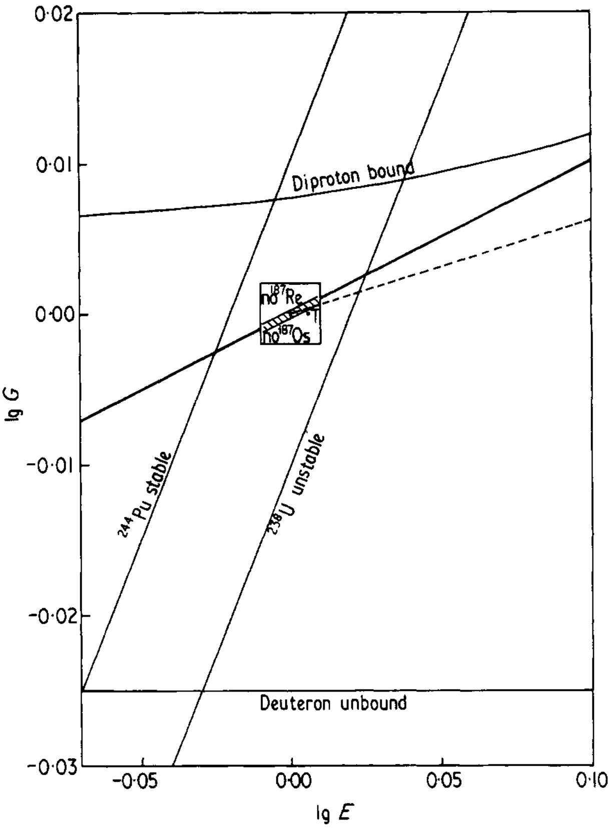

Figure 2. Constraints on nuclear structure as a function of strong and electromagnetic coupling.

the small quantities, we have $G \simeq$ E to produce constant Δ, and hence an unchanging halflife.

The thin band does not extend indefinitely, as it intersects the ^{238}U stability line in the region $E = G = 1{\cdot}005$ and also the diproton line shortly afterwards. In the opposite direction it intersects the ^{244}Pu stability line in the region $E = G = 0{\cdot}995$. The restriction of the allowable region to a finite area is thus a consequence of the different G, E functional dependence for α and β decay (and the relative independence of the diproton energy on E) so that these lines intersect.

In spite of Dyson's conclusion to the contrary it is felt that Teller's suggestion is still well within the limits discussed, if we include a modest decrease of G with time —for instance $G \sim E$ in the region $G = 1$. However, at an earlier epoch to avoid crossing the diproton line the G dependence would need to be somewhat slower, though greater than the limiting broken line. Because the diproton line curves upwards with increasing gradient, remembering the present accuracy, the broken curve need not cross the diproton line at any stage.

Although the limits given here appear to rule out Gamow's suggestion, the margin of error is rather small if we assume a $g_s^2 \propto e^2 \sim t$ time dependence, and especially as ^{244}Pu actually has been found naturally.

Acknowledgments

I should like to thank Dr W Saslaw and Dr B Carter for helpful discussions.

References

Bachall J N and Schmidt M 1967 *Phys. Rev. Lett.* **19** 1294–5
Broulik B and Trefil J S 1971 *Nature, Lond.* **232** 246–7
Dyson F J 1967 *Phys. Rev. Lett.* **19** 1291–3
—— 1971 *Scientific American* **225** 51–9
Gamow G 1967 *Phys. Rev. Lett.* **19** 759–61, 913–14
Hoffman D C, Lawrence F O, Mewherter J L and Rourke F M 1971 *Nature, Lond.* **234** 132–4
Okamoto L and Pask C 1971 *Ann. Phys.* **68** 18–56
Peres A 1967 *Phys. Rev. Lett.* **19** 1293–4
Teller E 1948 *Phys. Rev.* **73** 801–2