CERN-TH.4719/87

COULD A DILATON SOLVE THE COSMOLOGICAL CONSTANT PROBLEM ?

John Ellis, N.C. Tsamis and M. Voloshin^{*)}

CERN - Geneva

A B S T R A C T

Triggered by a solution to the cosmological constant problem recently proposed by Peccei, Solà and Wetterich (PSW), we compile astrophysical and cosmological constraints on a light scalar particle such as a pseudodilaton. The cosmological upper limit on the density of coherent waves in the present-day Universe and data from the Viking space probe together constrain the Compton wavelength of a scalar with gravitational strength couplings to be $\gtrsim 10^{10}$ light-years. Pseudodilaton of the PSW type that get their masses from quantum effects in QCD can only have couplings in the range allowed for invisible axions. A more natural possibility would be that quantum effects associated with other, stronger interactions (cosmocolour) generate the PSW pseudodilaton mass, allowing a larger range of coupling strengths. Alternatively, one could solve the cosmological constant problem by postulating a true dilaton in a finite theory with no intrinsic scale. Speculative possibilities for such a theory are mentioned.

^{*)} Permanent address: Institute for Theoretical and Experimental Physics, Moscow, USSR

The problem of the cosmological constant has been discussed in uncountably many papers, ever since it was realized that the fluctuations of quantum fields naturally produce a vacuum energy density which is many orders of magnitude larger than that allowed by cosmological observation. A new solution to this problem has recently been proposed by Peccei, Solà and Wetterich (PSW)¹⁾, which employs a pseudo-Goldstone boson of dilatation symmetry. This light scalar pseudodilaton particle has some affinities with the invisible axion which had previously been proposed as a solution to the strong CP problem²⁾. We have been triggered by PSW to compile the many astrophysical and cosmological constraints on light, weakly coupled scalar particles. These bounds apply not only to the pseudodilaton of PSW, but also to Brans-Dicke scalars, to conjectured scalars carrying the fifth force³⁾, and to any other light scalars (not yet) invented for whatever reason. After making this general-purpose compilation, we then re-examine the specific PSW proposal¹⁾, commenting on its cosmological difficulties, on some theoretical shortcomings, and suggesting acceptable variations on the basic idea of PSW. Before presenting our arguments in detail, we first summarize our principal conclusions.

The cosmological upper limit on the density of coherent waves in the present-day Universe⁴⁾ only allows a scalar boson with couplings to matter of gravitational strength or weaker if its Compton wavelength exceeds 10^2 light-years. When combined with information from the Viking space-probe⁵⁾, this limit is extended to 10^{10} light-years. In particular, this argument would rule out light scalars of the type proposed⁶⁾ to explain gravitational anomalies, the fifth force³⁾, etc. Important constraints on more strongly coupled light scalars are imposed by astrophysics, especially the energetics of red giants, white dwarfs⁷⁾ and the supernova SN 1987a⁸⁾. The allowed domains of mass/coupling parameter space for scalars ϕ include $m_\phi \lesssim 10^{-33}$ eV and $G_\phi \lesssim 10^{-3} G_N$, a small domain similar to that for the invisible axion⁴⁾, and a large domain⁹⁾ where $m_\phi \gtrsim 100$ keV and $G_N \ll G_\phi \lesssim G_F$. The specific proposal of PSW lies well inside the cosmologically disallowed region. We comment that the PSW mechanism works only if the vacuum energy density due to other physics (electroweak, etc.) is of similar magnitude to the QCD contribution¹⁰⁾. A more natural alternative would be to postulate a similar mechanism employing quantum effects associated with another gauge interaction (cosmocolour) which becomes strong at an energy scale $\Lambda_C \gtrsim 100$ GeV, resulting in less fine-tuning of the electroweak energy density and a larger pseudodilaton mass. This possibility would be allowed for a wide range of coupling strengths stronger than gravitational interactions. Alternatively, one could solve the cosmological constant problem by postulating a true dilaton in a finite theory with no intrinsic scale. Unsatisfactory examples of such

theories are $N = 4$ supersymmetric Yang-Mills theory¹¹⁾ or $N = 4$ conformal supergravity¹²⁾. Such speculations lead us beyond string theory, in both the energy scale and the number of internal dimensions.

After the introduction, we now discuss our compilation of bounds on scalar masses and couplings exhibited in the figure. Following PSW, we assume couplings to gauge bosons of the form $\phi F^{\mu\nu} F_{\mu\nu} / f_\phi$. In the general pseudodilaton case, we do not need to assume tree-level couplings to elementary fermions of the form $\phi \bar{f} f \cdot m_f / f_\phi$ which appear in many theories. Even in their absence, the ϕ -gluon-gluon coupling induces a non-zero coupling to hadronic matter which is discussed in Ref. 1). In the case of a true massless dilaton, the couplings are fixed by dilatation symmetry: $\phi T_\mu^\mu / f_\phi$, and the $\phi F^2 / f_\phi$ coupling is actually ineffective at low momentum transfers $q^2 \ll m_f^2$.

Starting from the right of the figure, the two diagonal lines labelled B/A exclude regions where $\phi \rightarrow \gamma\gamma$, etc. decays provide more than a factor $10^5/10$ of entropy, which are the amounts we deem acceptable before/after cosmological nucleosynthesis⁹⁾. Lower and to the left, the curve S excludes a region where red giants and white dwarf stars lose too much energy by ϕ emission induced by the Primakoff process⁷⁾. Its right-hand boundary marks the kinematic limit for ϕ emission from red giants with $T \sim 10^9$ °K $\sim 10^5$ eV. Just above is the curve SN bounding the region excluded by the upper limit on ϕ emission by the supernova SN 1987a⁸⁾. Its right-hand boundary is at $m_\phi \sim 2T \sim 10$ MeV, whilst its upper boundary is at a slightly larger value of f_ϕ than the red giant curve S. In the middle of the figure, the composite curve F bounds the region excluded by searches for intermediate-range forces¹³⁾. They only allow scalars ϕ above and to the right of the curve F. To the left of the figure, the curve V bounds the region excluded by measurements of signal delay from the Viking probe⁵⁾. These are useful because, unlike gravity, a dilaton field has no effect on the propagation of radiowaves. Again, the allowed region is above and to the right of the curve V. Finally, we come to the diagonal line C, above which the cosmological density of coherent ϕ waves would exceed the present closure density⁴⁾. In view of its importance, we describe this bound in more detail.

The equation of motion of a classical scalar field in our expanding Universe can be approximated by

$$\ddot{\phi} + 3H\dot{\phi} + m_\phi^2 \phi = 0 \quad (1)$$

where we have set the minimum of the potential at $\phi = 0$ and $H \sim (\rho/m_P^2)^{1/2}$ is the Hubble expansion rate. At an early epoch, when $H > m_\phi$, the damping term in (1) dominates, the ϕ field rolls very slowly down the potential and loses very little energy. In general, one expects the initial value of ϕ to be $O(f_\phi)$ away from the minimum, so that ϕ field energy

$$\rho_\phi \sim m_\phi^2 f_\phi^2 \quad (2)$$

during this epoch. If the Compton wavelength $L_\phi = 1/m_\phi$ of ϕ is longer than the size $r_0 \sim t_0 \sim H_0^{-1} \sim 10^{10}$ y of the observable Universe, we are still in this "early" epoch, and the energy density (2) should be less than the present upper limit given by the critical closure density $\rho_c \sim 10^{-46} (\text{GeV})^4$. It is easy to check that if the density is given by Eq. (2), then the "earliness" condition $H > m_\phi$ is respected only if $f_\phi > m_P$. Hence the part C_E of the coherent energy density line C.

If the Compton wavelength of ϕ is less than the size of the observable Universe, then the ϕ field has started to oscillate about its minimum with an amplitude $A(t)$. During these oscillations, the coherent energy density $\propto T^3$ like conventional non-relativistic matter:

$$\rho_\phi(T) = m_\phi^2 f_\phi^2 (T/T_i)^3 \quad (3)$$

where T_i is the temperature at the beginning of the oscillations which is when m_ϕ becomes $\sim H(T_i)$. If the effective temperature T falls to T_i during a matter-dominated epoch, Eq. (3) tells us that thereafter $\rho_\phi(T)$ is a fixed fraction of the total energy density. This is less than the closure density at epochs after $H(T_i) = m_\phi$ only if $f_\phi < m_P$. Hence the horizontal part C_N of the line C. Finally, if during a radiation-dominated epoch a phase transition at $T_c \sim \sqrt{f_\phi m_\phi}$ generates $m_\phi > H \sim T_c^2/m_P$ dynamically, then $T_i \sim T_c$ and the standard axion analysis based on Eq. (3) tells us that⁴⁾

$$m_\phi^{1/2} f_\phi^{3/2} \lesssim 10^{-26} m_P^2 \quad (4)$$

providing the part C_R of the line C. If, on the other hand, the mass m_ϕ generated by the phase transition is less than $H(T_c \sim \sqrt{f_\phi m_\phi})$, which is the case if $f_\phi > m_P$, then T_i is less than T_c and is simply given by $H(T_i) \sim T_i^2/m_P \sim m_\phi$. In this case Eq. (4) is replaced by

$$m_\phi^{1/2} f_\phi^2 \lesssim 10^{-26} m_P^{5/2}. \quad (4')$$

This case is indicated as a notch in the line C_N . We remind the reader that only the area below the curve C is cosmologically allowed.

Looking now at the overall picture presented by the figure, we see that there are three allowed regions of the (m_ϕ, f_ϕ) parameter space, indicated by somewhat thicker lines. The right-hand allowed region extends off to higher m_ϕ ⁹⁾. For $m_\phi \lesssim 100$ GeV, there are lower bounds $f_\phi \gtrsim 100$ GeV from particle physics which have not been indicated explicitly. In this region, the cosmological upper bound on f_ϕ does not allow $f_\phi \sim m_P$ unless $m_\phi \gtrsim 10^{12}$ GeV. The central allowed region

$$10^{10} \text{ GeV} \lesssim f_\phi \lesssim 10^{12} \text{ GeV}, \quad 10^{-4} \text{ eV} \lesssim m_\phi \lesssim 1 \text{ eV} \quad (5)$$

corresponds roughly to that allowed for the invisible axion⁴⁾, although it is somewhat reduced by laboratory searches for bulk forces. The left-hand allowed region is

$$f_\phi \gtrsim 10^{20} \text{ GeV}, \quad m_\phi \lesssim 10^{-33} \text{ eV} \quad (6)$$

and we will return to it later.

It is clear from the figure that intermediate range scalars with couplings weaker than gravity, of the types postulated⁶⁾ to explain gravitational anomalies, the fifth force, etc., are totally excluded by the cosmological energy density argument. Included among these particles is the PSW pseudodilaton with

$$m_\phi \sim 10 \Lambda_{\text{QCD}}^2 / f_\phi, \quad f_\phi \gtrsim 10 m_P \quad (7)$$

which triggered this study. We will now discuss whether the PSW idea can be rescued in some form.

The proposal¹⁾ of PSW is that the electroweak and other possible shorter-range interactions are scale-invariant at the classical level. This is possible if scale invariance is spontaneously broken, there is a dilaton S with decay constant f_S , and all dimensional parameters in the Lagrangian appear multiplied by the corresponding powers of $\exp(S/f_S)$. For example, the Higgs squared mass parameter $\mu^2 \rightarrow \mu^2 \exp(2S/f_S)$, and any field-independent energy density term $\Lambda_0 \rightarrow \Lambda_0 \exp(4S/f_S)$, so that the conventional Higgs potential becomes

$$V_{EW}(H, S) = -\mu^2 e^{2S/f_S} |H|^2 + \frac{\lambda}{2} (|H|^2)^2 + \Lambda_0 e^{4S/f_S} \quad (8)$$

The Higgs vacuum expectation value is

$$\langle 0|H|0 \rangle = \left[\frac{\mu^2}{\lambda} e^{2S/f_S} \right]^{1/2} \quad (9)$$

and the resulting vacuum energy is

$$V_{EW}(S) = \left(\Lambda_0 - \frac{\mu^4}{2\lambda} \right) e^{4S/f_S} \quad (10)$$

The as yet undetermined value of S is supposed to be fixed by quantum corrections in an exact gauge theory such as QCD, which make a contribution to the potential

$$\Delta V_{QCD} = \theta_\mu^\mu S / f_S \quad (11)$$

where $\theta_\mu^\mu \equiv \langle 0 | \beta_{QCD}(\alpha_S) / (2\alpha_S) F_{\mu\nu} F^{\mu\nu} | 0 \rangle$ is the v.e.v. of the trace of the QCD energy-momentum tensor¹⁴⁾. The condition for minimizing the full potential ($V_{EW} + \Delta V_{QCD}$) with respect to S is

$$0 = \frac{\partial V_{EW}}{\partial S} + \frac{\theta_\mu^\mu}{f_S} = \frac{1}{f_S} \left[4 \left(\Lambda_0 - \frac{\mu^4}{2\lambda} \right) e^{4S/f_S} + \theta_\mu^\mu \right] = 0 \quad (12)$$

The key observation of PSW is that this is also the condition for the vanishing of the cosmological constant

$$T_\mu^\mu = f_S \frac{\partial V_{EW}}{\partial S} + \theta_\mu^\mu = 0 \quad (13)$$

Hence this combination of classical scale invariance for the electroweak interactions broken by quantum effects in QCD provides a mechanism for dynamical relaxation of the cosmological constant, similar to the axion mechanism²⁾ for strong CP invariance.

As was pointed out by PSW, this mechanism requires V_{EW} and ΔV_{QCD} to have opposite signs. The phenomenologically-determined¹⁰⁾ negative sign of θ_μ^μ tells us that ΔV_{QCD} is in fact negative. This choice is actually necessary for the viability of the PSW mechanism, since vacuum stability requires $\partial^2 V / \partial S^2 > 0$ at the minimum which is only true if $V_{EW} > 0$ and hence ΔV_{QCD} is negative. We would

also like to point out that it is not completely obvious that $\theta_\mu^\mu \propto \Lambda_{\text{QCD}}^4$ is independent of S . It is at least possible that QCD is embedded in a larger, spontaneously broken theory at a large scale m_X in such a way that $\Lambda_{\text{QCD}} \propto m_X$, whilst m_X contains an $\exp(S/f_S)$ dependence analogous to the Higgs mass parameter in $V_{\text{EW}}(10)$. Although this S -dependence could be compensated by the renormalization group, this point should be checked in any fully unified theory embodying the PSW mechanism.

It is important to note that the vacuum energy cancellation (13) tells us that

$$V_{\text{EW}} = \frac{1}{4} f_S \frac{\partial V_{\text{EW}}}{\partial S} = -\frac{1}{4} \theta_\mu^\mu = O(\Lambda_{\text{QCD}}^4) \quad (14)$$

at the extremum of the potential. Therefore the classical vacuum energy $\Lambda_0 e^{4S/f_S}$ present before electroweak symmetry breakdown must be prepared equal and opposite to $\mu^4/2\lambda e^{4S/f_S} = \langle 0|H|0\rangle^4/2$, apart from a small correction of order $(\Lambda_{\text{QCD}}/\langle 0|H|0\rangle)^4 \sim 10^{-12}$. Although disconcerting, this fine-tuning is of course much less severe than the fine-tuning of 120 decimal places that was necessary before. Nevertheless, this lesser fine-tuning can be avoided by postulating the presence of another set of exact gauge interactions analogous to QCD which become strong at a larger scale $\Lambda_c \sim (0.1 \text{ to } 1) \text{ TeV}$. We call these new strong interactions cosmocolour: we will not enter here into the question whether cosmocolour can be identified with any other type of new interaction proposed in recent years, such as technicolour, metacolour, hypercolour, supercolour, etc. The PSW mechanism works just as well with $\theta_\mu^\mu(\text{cosmocolour})$ in Eqs. (11) et seq., and the QCD θ_μ^μ can be regarded as a small perturbation to it.

We now ask whether the PSW mechanism can be reconciled with cosmology, either in its original QCD form or in this cosmocolour version. In the case of QCD PSW, this is only possible if f_S is in the narrow range

$$f_S \sim (10^{10} \text{ to } 10^{11}) \text{ GeV} \quad (15)$$

where the dotted line in the figure crosses the central allowed region. There is a larger range of acceptable values for f_S in the cosmocolour version, corresponding to the dashed lines crossing the rightmost allowed region in the figure:

$$f_s \sim (10^2 \text{ to } 10^9) \text{ GeV} . \quad (16)$$

Although the PSW mechanism is tenable in these two cases, it is not attractive to contemplate values of f_s so much smaller than m_p .

Thus we contemplate finally the possibility of solving the problem of the cosmological constant with a dilaton in the leftmost allowed region of the figure. This could either be a very slightly pseudodilaton with $m_s \lesssim 10^{-33}$ eV and $f_s \gtrsim 10^{20}$ GeV as in (6), or else a true dilaton with $m_s = 0$. A true dilaton could only appear in a finite theory with no intrinsic scale at all, in which all scales are generated spontaneously. Two examples of such theories are known, namely $N = 4$ ^{*)} supersymmetric Yang-Mills ¹¹⁾ and $N = 4$ conformal supergravity ¹²⁾. However, neither of these seems very realistic: $N = 4$ Yang-Mills does not contain gravity, and conformal supergravity suffers from all the usual unitarity and/or negative energy problems associated with R^2 gravity theories ¹⁶⁾. String theory is an example of what we do not want for this dilaton solution of the cosmological constant problem: although it is presumably finite, it has an intrinsic scale provided by the string tension ^{**)}. It is amusing to ask whether there could be some even more fundamental theory beyond string which could have the properties we seek. All quantum theories involve sums over phases $\beta: \Sigma e^{i\beta}$ where β is the action. The first-quantized action for a simple point particle is proportional to the length L of its world line: $\beta \propto \int dL$, whilst for string it is proportional to the area of the worldsheet: $\beta \propto \int dA$. Since both lengths and areas have intrinsic dimensions, their actions must contain dimensionful parameters, namely the string tension σ in the case of string:

$$\sum e^{i\beta} : \quad \beta = \sigma \int dA . \quad (17)$$

One could imagine extrapolating this philosophy to extended objects in arbitrary dimensions:

$$\beta = [M]^n \int dV^n \quad (18)$$

*) And some $N = 2$ supersymmetric Yang-Mills models ¹⁵⁾.

***) If string theory does solve the problem of the cosmological constant, presumably it does so via a different mechanism.

and ask when the dimensionful parameter is absent. Clearly, this is the case only when $n = 0$, in which case the theory is a theory of points and pure numbers. One could either interpret this conclusion as a reductio ad absurdum of the pure dilaton idea, or else speculate that all of physics, including perhaps string theory, is merely a low-energy manifestation of a theory of numbers which becomes manifest at some spontaneously generated scale above 10^{20} GeV. Maybe we had better stop this paper at this point.

ACKNOWLEDGEMENTS

One of us (M.V.) thanks the CERN Theory Division for its kind hospitality while this work was done.

REFERENCES

- 1) R.D. Peccei, J. Solà and C. Wetterich - DESY Preprint DESY 87/026 (1987).
- 2) R.D. Peccei and H.R. Quinn - Phys.Rev.Lett. 38 (1977) 1440 and Phys.Rev. D16 (1977) 1791.
- 3) E. Fischbach et al. - Phys.Rev.Lett. 56 (1986) 3.
- 4) L. Abbott and P. Sikivie - Phys.Lett. 120B (1983) 133;
J. Preskill, M. Wise and F. Wilczek - Phys.Lett. 120B (1983) 127;
M. Dine and W. Fischler - Phys.Lett. 120B (1983) 137.
- 5) I.I. Shapiro et al. - J.Geophys.Res. 82 (1977) 4329.
- 6) D. Chang, R.N. Mohapatra and S. Nussinov - Phys.Rev.Lett. 55 (1985) 2835.
- 7) M.I. Vysotski et al. - JETP 27 (1978) 533;
K. Sato - Progr.Theor.Phys. 60 (1978) 1942;
D.A. Dicus et al. - Phys.Rev. D15 (1978) 1829;
D.S.P. Dearborn, D.N. Schramm and G. Steigman - Phys.Rev.Lett. 56 (1986) 26.
- 8) J. Ellis and K.A. Olive - CERN Preprint TH. 4701/87 (1987).
- 9) J. Ellis, D.V. Nanopoulos and M. Quiros - Phys.Lett. 174B (1986) 176.
- 10) M.A. Shifman, A.I. Vainshtein and V.I. Zakharov - Nucl.Phys. B147 (1979) 385.
- 11) S. Mandelstam - J.Phys. 43 (1982) C3-331;
L. Brink, O. Lindgren and B. Nilsson - Nucl.Phys. B212 (1983) 401 and Phys.Lett. 123B (1983) 328.
- 12) E.S. Fradkin and A.A. Tseytlin - Phys.Lett. 134B (1984) 307.
- 13) A. De Rujula - Phys.Lett. 180B (1986) 213.
- 14) R. Crewther - Phys.Rev.Lett. 28 (1972) 1421;
M.S. Chanowitz and J. Ellis - Phys.Lett. 40B (1972) 397;
J.C. Collins - Phys.Rev. D14 (1976) 1965;
S.L. Adler, J.C. Collins and A. Duncan - Phys.Rev. D15 (1977) 1712;
J.C. Collins, A. Duncan and S.D. Joglekar - Phys.Rev. D16 (1977) 438.
- 15) For a review, see: P. West - "Supersymmetry - a Decade of Development", ed. P. West (Adam Hilger, Bristol, 1986), p. 69.
- 16) K. Stelle - Phys.Rev. D16 (1977) 993.

FIGURE CAPTION

The plane of scalar masses m_ϕ and coupling parameters f_ϕ , showing the domains excluded by the different astrophysical and cosmological arguments described in the text. The lines A, B indicate bounds on entropy release in $\phi \rightarrow \gamma\gamma$ decays, and are taken from Ref. 9). The lines S, SN come from energy losses in stars⁷⁾ and the supernova SN 1987a⁸⁾. The line F is a compilation of constraints on intermediate range forces taken from Ref. 13), and the constraint V is provided by the Viking probe⁵⁾. The line $C = C_E + C_N + C_R$ is the cosmological bound on the coherent energy density⁴⁾. The disallowed side of each line is indicated by hatching. The boundaries of the allowed regions are indicated by thick lines. The diagonal dotted (dashed) lines show values of (m_ϕ, f_ϕ) suggested by values of Θ_μ^μ in QCD¹⁰⁾ (cosmocolour) models. The PSW pseudodilaton¹⁾ is in the disallowed region above the line C.