The Strong CP Problem and Its Implications

PITP , 2017

Michael Dine

Department of Physics
University of California, Santa Cruz.

July, 2017

Lecture 1: The Strong CP Problem
Lecture 2: Solutions of the Strong CP Problem: An Assessment
Lecture 3: Axion Cosmology and Axion Searches: Old and New Ideas
The lectures will be self contained. But some suggested reading:
Suggested Reading: Supersymmetry and String Theory: Beyond the Standard Model by M. Dine, sections 5.4, 5.5, 19.2
There are many recent reviews. Some examples:
J. Kim:
https://arxiv.org/abs/1703.03114
D. Marsh:
https://arxiv.org/abs/1510.07633
Kawasaki et al:
https://arxiv.org/abs/1301.1123
P. Sikivie:
https://arxiv.org/abs/astro-ph/0610440
Helpful to do some review of anomalies. Choose your favorite textbook (Peskin and Schroder, Schwarz, Dine)

Lecture 1: The Strong CP Problem

1. Our understanding of QCD: Lagrangian parameters, symmetries
2. Anomalies; CP violating parameters, U(1) problem and their connections
3. Instantons as an indicator of θ dependence and consequences of the anomaly
4. Consequences of θ-dependence: the Neutron electric Dipole moment.
5. Possible Solutions

Our present understanding of the Strong Interactions

QCD: very simple lagrangian, governed by symmetries and particle content:

$$\mathcal{L} = -\frac{1}{4g^2}(F^a_{\mu\nu})^2 + \sum_f \bar{q}_f(i\not{D} - m_f)q_f \quad (1)$$

Here $F_{\mu\nu}$ is the field strength of QCD,q_f the various quark flavors.

Extremely successful.

- At very high energies, detailed, precise predictions, well verified at Tevatron and LHC (dramatically in the context of Higgs discovery.
- At low energies, precision studies with lattice gauge theory.

QCD Parameters from Lattice Gauge Theory

The parameters of the theory: α_s at some scale; quark masses. Current results from lattice simulations (summarized by the FLAG working group)

$$m_u = 2.16\ (9)(7)\mathrm{MeV} \quad m_d = 4.68\ (14)(7)\mathrm{MeV}\, m_s = 93.5(2.5)\mathrm{MeV} \tag{2}$$

(Numbers are in $\overline{MS}$ scheme at 2 GeV.)

$\alpha_s(M_Z) = 0.1192 \pm 0.0011$ from lattice gauge theory; compatible with other measurements.

Would seem we know all we could want to know

Symmetries of QCD

The u, d, and to a lesser extent the s quarks are light. Working with left-handed fields,

$$q = \begin{pmatrix} u \\ d \\ s \end{pmatrix} \quad \bar{q} = \begin{pmatrix} \bar{u} \\ \bar{d} \\ \bar{s} \end{pmatrix} \tag{3}$$

For zero quark mass, the theory has symmetry $U(3) \times U(3)$:

$$q \to Uq; \quad \bar{q} \to V\bar{q} \tag{4}$$

The symmetry is spontaneously broken to a vector subgroup, $SU(3) \times U(1)_B$:

$$\langle \bar{q}_{\bar{f}} q_f \rangle = a\Lambda^3 \delta_{\bar{f}f} \tag{5}$$

Goldstone bosons: π, K, η (eight particles). These particles behave as expected for GB's ("current algebra").

Where's the ninth? The η' is much heavier than the others. E.g.

$$m'_\eta = 958 \text{ MeV}; \; m_\eta = 548 \text{ MeV} \tag{6}$$

One feature of the axial $U(1)$ current which might account for this: it is not strictly conserved. $q \to e^{i\alpha q} \quad \bar{q} \to e^{i\alpha q}$ In four component language,

$$j_5^\mu = \bar{q}\gamma^\mu\gamma^5 q; \quad \partial_\mu j_5^\mu = \frac{N_f}{16\pi^2} F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{7}$$

Here $\tilde{F}$ (“F-dual") is

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}; \quad F\tilde{F} = 2\vec{E}\cdot\vec{B}. \tag{8}$$

The right hand side is a total derivative, $\frac{1}{16\pi^2}\partial_\mu K^\mu$

$$K_\mu = \epsilon_{\mu\nu\rho\sigma}\left(A^a_\nu F^a_{\rho\sigma} - \frac{2}{3} f^{abc} A^a_\nu A^b_\rho A^c_\sigma\right) \tag{9}$$

So one can define a new conserved current,

$$\tilde{j}^\mu = j^\mu_5 - \frac{N_f}{16\pi^2} K^\mu \tag{10}$$

But this current is not gauge invariant, so its status requires investigation. We'll say much more about it shortly, but the charge associated with this current is not conserved. Configurations with $A^\mu \sim \frac{1}{r}$ are already important in a semi-classical treatment. So there is no mystery in the absence of a ninth Goldstone boson.

Strong CP Problem

But the assertion that $F\tilde{F}$ solves the $U(1)$ problem raises a puzzle: one can now add to the QCD lagrangian

$$\mathcal{L}_\theta = \frac{\theta}{16\pi^2} F\tilde{F}. \tag{11}$$

This term violates *parity*, and thus CP.

Again, while

$$F\tilde{F} = \partial_\mu K^\mu; \tag{12}$$

K^μ is not gauge invariant. If there are important configurations, say, in the path integral, for which the A^3 term falls off as $1/r^3$, then we can't drop the surface term. Examples of such configurations will be discussed shortly (instantons).

The possibility of a non-zero θ is related to another possible source of *CP* violation. In writing the QCD Lagrangian, we did not commit to two or four component fermions. In the language of two component fermions, we might have taken complex masses. With four component fermions, we might have included $\bar{q}\gamma_5 q$ terms in the quark lagrangian.

One might have said: what's the big deal? If we had written (two component form)

$$\bar{q}_f m_{\bar{f}f} q_f + \mathrm{c.c.} \tag{13}$$

we could, by separate $U(3)$ transformations of the q and $\bar{q}$ fields, have rendered m (the quark mass matrix) diagonal and real.

But we have just asserted that this redefinition has an anomaly. Indeed, if we think about the effect of such a transformation on the lagrangian,

$$\delta\mathcal{L} = \alpha\partial_\mu j_5^\mu = \alpha\frac{1}{16\pi^2}F\tilde{F}. \tag{14}$$

So we can trade a phase in the quark mass matrix for θ, or vice versa. The invariant quantity is often called $\bar{\theta}$,

$$\bar{\theta} = \theta - \arg\ \det m_q. \tag{15}$$

Anomaly in the current $j^\mu = \bar{q}\gamma^\mu\gamma^5 q$: A calculation

We are interested in a mass term, $me^{i\alpha}\bar{q}q$ There are many ways to think about the anomaly and I cannot do justice to the subject in ten minutes. But I can at least indicate how this effect comes about. In four component language, we have a coupling, for small α:

$$m\,\alpha\bar{q}\gamma^5 q \tag{16}$$

Let's replace m by a field (this might be one of the pseudoscalar mesons, or, as we will see later, the axion). This allows some momentum flow through the diagram.

It is convenient to work with the graph in four component notation since the fermion is massive. The basic expression has the form, after introducing Feynman parameters:

$$N_f m(q)\alpha \int\int dx_1 dx_2 f(x_1, x_2) \frac{d^4k}{(2\pi)^4[k^2+m^2]^3} \mathrm{Tr}\gamma_5 \tag{17}$$

$$(\not{p}_1 \not{A}(p_1) ptwoslash \not{A}(p_2))$$

$$= \frac{1}{16\pi^2}\alpha\epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu A(p_1)^\rho A(p_2)^\sigma.$$

This is the anticipated result.

Exercise: Verify the expression above

Some things to note:

1. If m is a constant, the result vanishes ($p_1 = -p_2$). This is related to the total derivative, and the rather trivial (plane wave) nature of the external fields.
2. If we did the computation using the background field formalism, we would obtain, also, $\partial_\mu j^\mu = \frac{N_f}{16\pi^2} F\tilde{F}$. This result would hold for non-trivial backgrounds (magnetic monopoles in a $U(1)$ theory, instantons in the non-abelian theory) for which $\int d^4x F\tilde{F}$ is non-zero.
3. It is interesting that $\alpha F\tilde{F}$ term appears in the effective action. Surprising since would seem one could just rotate away.

Instantons

QCD is a strongly coupled theory, and providing a reliable answer to this question requires strong coupling methods (lattice gauge theory). But a semiclassical analysis, while not reliable, indicates that one cannot neglect $\int d^4x \partial_\mu K^\mu$. As a result, the axial current is not conserved and θ is physical.

In the Euclidean functional integral

$$Z = \int [dA][dq][d\bar{q}]e^{-S} \tag{18}$$

it is natural to look for stationary points of the effective action, i.e. finite action, classical solutions of the theory in imaginary time. These *instanton* solutions can be found rather easily. The following tricks simplify the construction, and turn out to yield the general solution. First, note that the Yang–Mills action satisfies an inequality, the Bogomolnyi bound:

$$\int (F \pm \tilde{F})^2 = \int (F^2 + \tilde{F}^2 \pm 2F\tilde{F}) = \int (2F^2 \pm 2F\tilde{F}) \geq 0. \tag{19}$$

So the action is bounded by $|\int F\tilde{F}|$, with the bound being saturated when

$$F = \pm\tilde{F} \tag{20}$$

i.e. if the gauge field is (anti-) self-dual.

This is a first order differential equation. Comparatively easy to solve.

The Instanton Solution

't Hooft presented the instanton in a fashion which is useful for actual computations. Defining the symbol η:

$$\eta_{aij} = \epsilon_{aij};\ \eta_{a4i} = -\eta_{ai4} = -\delta_{ai};\quad \bar{\eta}_{a\mu\nu} = (-1)^{\delta_{a\mu}+\delta_{a\nu}}\eta_{a\mu\nu} \tag{21}$$

the instanton takes the simple form:

$$A^a_\mu = \frac{2\eta_{a\mu\nu}x^\nu}{x^2+\rho^2} \tag{22}$$

while the field strength is given by:

$$F^a_{\mu\nu} = \frac{4\eta_{a\mu\nu}\rho^2}{(x^2+\rho^2)^2} \tag{23}$$

That this configuration solves the equations of motion follows from:

$$\eta_{a\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\eta_{a\alpha\beta}. \tag{24}$$

so $F = \tilde{F}$.
The η symbols are connected to the embedding of $SU(2)$ of the gauge group in an SU(2) subgroup of $O(4) = SU(2) \times SU(2)$. This can be understood by noting:

$$\eta_{a\mu\nu} = \frac{1}{2}\mathrm{Tr}(\sigma^a \sigma_{\mu\nu}) \quad \bar{\eta} = \mathrm{Tr}(\sigma^a \bar{\sigma}_{\mu\nu}). \tag{25}$$

Since $F = \tilde{F}$, the equations of motion are satisfied. Note the $1/r$ falloff of A^μ, as opposed to the $1/r^4$ falloff of $F_{\mu\nu}$.

Topological Charge

Asymptotically, A^μ is a pure gauge. The gauge transformation maps the three sphere onto the gauge group. $\frac{1}{16\pi^2}\int d^4x F\tilde{F}$ measures the number of times that the sphere is mapped into the gauge group.

$$\frac{1}{16\pi^2}\int d^4x F\tilde{F} = \frac{1}{16\pi^2}\int d^4x \partial_\mu K^\mu = 1 \qquad (26)$$

At large x,

$$A_\mu = \frac{\eta_{a\mu\nu} x^\nu \tau^a}{x^2} \tag{27}$$

$$= i g^{-1} \partial_\mu g$$

[**Exercise:** show what $g = \frac{x_4 + i\vec{x}\cdot\vec{\tau}}{r}$].

Consequences of the Instanton

Having found a classical solution, we want to integrate about small fluctuations about it. Including the θ term, these have the form

$$\langle \mathcal{O} \rangle = e^{-\frac{8\pi^2}{g^2}} e^{i\theta} \int [d\delta A][dq][d\bar{q}] \exp\left(-\frac{\delta^2 S}{\delta A^2} \delta A^2 - S_{q,\bar{q}} \right) \mathcal{O}. \quad (28)$$

Now S contains an explicit factor of $1/g^2$. As a result, the fluctuations are formally suppressed by g^2 relative to the leading contribution. The one-loop functional integral yields a product of determinants for the fermions, and of inverse square root determinants for the bosons.

Both the bosonic and fermionic quadratic fluctuation operators have zero eigenvalues. For the bosons, these potentially give infinite contributions to the functional integral, and they must be treated separately. The difficulty is that among the variations of the fields are symmetry transformations: changes in the location of the instanton (translations), rotations of the instanton, and scale transformations.

More explicitly

$$A_\mu^a(x) = \Omega \frac{2\eta_{a\mu\nu}(x-x_0)^\nu}{x^2+\rho^2} \tag{29}$$

where Ω denotes a global gauge transformation (or rotation).

Consider translations. For every solution, there is an infinite set of solutions obtained by shifting the origin (varying x_0). Instead of integrating over a coefficient, c_0, we integrate over the *collective coordinate* x_0 (one must also include a suitable Jacobian factor). The effect of this is to restore translational invariance in Green's functions. Similarly, the instanton breaks the rotational invariance of the theory. Correspondingly, we can find a three-parameter set of solutions and zero modes. Integrating over these rotational collective coordinates restores rotational invariance. (The instanton also breaks a global gauge symmetry, but a combination of rotations and gauge transformations is preserved.)

Finally, the classical theory is scale invariant; this is the origin of the parameter ρ in the solution. Again, one must treat ρ as a collective coordinate, and integrate over ρ. There is a power of ρ arising from the Jacobian, which can be determined on dimensional grounds. If d_O is the dimension of the operator $\mathcal{O}$, then, on dimensional grounds, one expects for the ρ dependence:

$$\int d\rho \rho^{-d_0-1}. \tag{30}$$

However, there is additional ρ-dependence because the quantum theory violates the scale symmetry. This can be understood by replacing $g^2 \to g^2(\rho)$ in the functional integral, and using

$$e^{-\frac{8\pi^2}{g^2(\rho)}} \approx (\rho M)^{b_0} \qquad (31)$$

for small ρ. For three-flavor QCD, for example, $b_0 = 9$, the leading operator has dimension 9, and the ρ integral diverges logarithmically for large ρ. This is just the statement that the integral is dominated by the infrared, where the QCD coupling becomes strong.

Fermion functional integrals introduce a new feature. In four-component language, it is necessary to treat q and $\bar{q}$ as independent fields. (In two-component language, this corresponds to treating q and q^* as independent fields.) So at one-loop order, we need to study:

$$\not{D} q_n = \lambda_n q_n \quad \not{D} \bar{q}_n = \lambda_n \bar{q}_n \tag{32}$$

$$q(x) = \sum a_n q_n(x), \quad (33)$$
$$S = \sum \lambda_n a_n^* a_n. \quad (34)$$

Then

$$\int [dq][d\bar{q}] e^{-S} = \prod_{n=0}^{\infty} da_n da_n^* e^{-\sum_{n \neq 0} \lambda_n a_n^* a_n}. \quad (35)$$

Zero eigenvalues of the Dirac operator are special. Because the zero modes do not contribute to the action, many Green functions vanish. For example, $\langle 1 \rangle = 0$. In order to obtain a non-vanishing result, we need enough insertions of q to “soak up" all of the zero modes.

The explicit form of the zero modes is not complicated. For $SU(2)$, for simplicity:

$$\not{D} q = 0 \quad \not{D} \bar{q} = 0 \tag{36}$$

and

$$q_0 = \frac{\rho}{(\rho^2 + (x - x_0)^2)^{3/2}} \zeta, \tag{37}$$

where ζ is a constant spinor.

q here might be $u, \bar{u}, d, \bar{d}$, etc.

We can put all of this together to evaluate a Green function which violates the classical $U(1)$ symmetry of the massless theory, $\langle \bar{u}(x)u(x)\bar{d}(x)d(x)\bar{s}(x)s(x)\rangle$. There is one zero mode for each of $u, d, s, \bar{u}, \bar{d}, \bar{s}$. The fields in this Green's function can soak up all of these zero modes. The effect of the integration over x_0 is to give a result independent of x, since the zero modes are functions only of $x - x_0$. The integration over the rotational zero modes gives a non-zero result only if the Lorentz indices are contracted in a rotationally invariant manner (the same applies to the gauge indices). The integration over the instanton scale size – the conformal collective coordinate – is more problematic, exhibiting precisely the infrared divergence we discussed earlier.

So we have provided some evidence that the $U(1)$ problem is solved in QCD, but no reliable calculation. What about θ-dependence? Let us ask first about θ-dependence of the vacuum energy. In order to get a non-zero result, we need to allow that the quarks are massive. Treating the mass as a perturbation, noting the path integral contains a term $e^{i\theta \frac{1}{16\pi^2}\int d^4x F\tilde{F}} = e^{i\theta}$ we obtain a result of the form:

$$E(\theta) = C\Lambda_{\mathrm{QCD}}^9 m_u m_d m_s \cos(\theta) \int d\rho \rho^7. \tag{38}$$

So we have evidence for θ-dependence, but again cannot do a reliable calculation. That we cannot do a calculation should not be a surprise. There is no small parameter in QCD to use as an expansion parameter. Fortunately, we can use other facts which we know about the strong interactions to get a better handle on both the $U(1)$ problem and the question of θ-dependence.

Neutron Electric Dipole Moment

A particularly sensitive test of CP conservation in the strong interactions is provided by the neutron electric dipole moment. This corresponds to an operator $\bar{n}\gamma_5\sigma_{\mu\nu}nF^{\mu\nu}$ (here F is the gauge field of electrodynamics). We might first guess that d_n is of order $e\theta \ \mathrm{Fm} \approx \theta 10^{-13} cm$. Then we might expect suppression by powers of quark mass. Our instanton analysis would suggest three factors of quark mass (m_q/m_n, say). We'll see in a moment that there is only one such factor; $m_u/m_n \sim 0.002$, so from the limit on $d_n < 10^{-26}\ \mathrm{e\ cm}$, we would have $\theta < 10^{-10}$. We'll give a sharper estimate now.

Consider, first, the coupling of pions to nucleons (for simplicity we'll consider a limit of approximate $SU(2) \times SU(2)$ symmetry, i.e. we'll just treat the u and d quarks as light. Also for simplicity I'll take the u and d quark masses identical. I'll quote general formulas in the end, and leave the derivations to you (or to an examination of the literature).

$$\mathcal{L}_{\pi NN} = \vec{\pi} \cdot \bar{N}(\vec{\tau} i \gamma_5 g_{\pi NN} + \bar{g}_{\pi NN})N. \tag{39}$$

The second term is CP violating. Its effects are directly measurable, in principle, but we will take this to a far more sensitive test in a moment.

As an aside, we can ask what this lagrangian means. We are used to the notion of an effective action for Goldstone bosons. This makes sense; the Goldstone bosons are the light fields, and we can obtain their lagrangian by integrating out heavy fields, such as nucleons and vector mesons. But in QCD, because baryon number is conserved, we can consider a sector with a fixed, non-zero baryon number. Baryon number one is the simplest. For low momenta, we can treat the nucleons as non-relativistic and ignore nucleon-anti-nucleon pairs, and ask about pion scattering amplitudes. This subject was developed extensively some time ago by Weinberg.

For example, in the baryon number one sector, we can study low momentum pion-nucleon processes by thinking of matrix elements such as $\langle \pi\pi \dots | \bar{q}q | \pi\pi \dots \rangle$ in terms of a background pion field, and obtain the amplitude from the non-linear chiral lagrangian.

The CP violating term in the underlying lagrangian is obtained from the quark mass term, $m\bar{q}q + \mathrm{c.c.}$, and performing the transformation $q \rightarrow e^{i\frac{\vec{\pi}\cdot\vec{\tau}}{2f_\pi}} q$:

$$\delta\mathcal{L} = \frac{m_q\theta}{2}\bar{q}\frac{\vec{\pi}}{f_\pi}\cdot\vec{\tau}q \tag{40}$$

(compare the chiral lagrangian, $\bar{q}mq \rightarrow \mathrm{Tr}mU$, $U = e^{i\frac{\vec{\pi}}{2f_\pi}}$, now $\bar{N}UN$. We need the matrix element between an initial and final nucleon state:

$$\frac{\theta m_q}{2f_\pi}\langle N_f|\bar{q}\tau^a q|N_i\rangle \tag{41}$$

The matrix element can be obtained from standard $SU(3)$ global symmetry (Gell-Mann) arguments.

Now this coupling induces a neutron electric dipole moment. The diagram is infrared divergent as $m_\pi \to 0$, and this term is readily extracted.

$$d_n = g_{\pi NN} \bar{g}_{\pi NN} \frac{\log(M_N/m_\pi)}{4\pi^2 M_N} \tag{42}$$

Working through the details:

$$d_n = g_{\pi NN} \frac{\theta m_u m_d}{f_\pi (m_u + m_d)} \langle N_f | \bar{q} \tau^a q | N_f \rangle \ln(m_p/m_\pi) \frac{1}{4\pi^2 m_p} \tag{43}$$

$$= 5.2 \times 10^{-16} \theta cm$$

(this is calculated in an approximation which becomes more and more reliable as the masses of the light quarks become smaller).

From the experimental limit, $d_n < 3 \times 10^{-26}\ \mathrm{e\ cm}$, one has $\theta < 10^{-10}$.

The Strong CP Problem

This is a puzzle. Why such a small dimensionless number?

$\theta \to 0$: strong interactions preserve CP. If not for the fact that the rest of the SM violates CP, would be *natural*.

Among naturalness problems, the strong CP problem is special in that it is of almost no consequence. We don't have to invoke anthropic selection to realize that if the cosmological constant was a few orders of magnitude larger than observed, the universe would be dramatically different. The same is true for the value of the weak scale and of the light quark and lepton masses. But if θ were, say, 10^{-3}, nuclear physics would hardly be different than we observe, since effects of θ are shielded by small quark masses.

Possible Resolutions

1. $m_u = 0$ If true, $u \to e^{-i\frac{\theta}{2}\gamma_5}u$ eliminates θ from the lagrangian. An *effective* m_u might be generated from non-perturbative effects in the theory (Georgi, McArthur; Kaplan, Manohar) Could result as an accident of discrete flavor symmetries (Banks, Nir, Seiberg), or a result of “anomalous" discrete symmetries as in string theory (M.D.)
2. CP exact microscopically, $\theta = 0$; spontaneous breaking gives the CKM phase but leads, under suitable conditions, to small effective θ (Nelson, Barr). In critical string theories, CP is an exact (gauge) symmetry, spontaneously broken at generic points in typical moduli spaces. A plausible framework.
3. A new, light particle called the axion dynamically cancels off θ.

Lecture 2: Solutions of the Strong CP Problem: An Assessment

1. $m_u = 0$: how plausible as an idea? Confrontation with lattice gauge theory
2. Small radiative corrections to θ in Standard Model. Possible realization through spontaneous CP violation (Barr-Nelson mechanism). Models. Virtues and problems.
3. Axions. Basic ideas. Virtues, problems, constraints.
4. Basics of Axion Physics
5. Astrophysical Constraints on Axions

Problems with each of these solutions:

1. $m_u = 0$. Lattice computations seem to rule out (the required non-perturbative effects do not seem to be large enough).
2. Spontaneous CP: special properties required to avoid large θ once CP is spontaneously broken. What would single out such theories?
3. Axions: promise and limitations.

$m_u = 0$

If $m_u = 0$, one can rotate away θ. More precisely, one requires, since $d_n \propto \theta m_u$ in this limit,

$$\frac{m_u}{\Lambda_{QCD}} < 10^{-10} \tag{44}$$

at the scale Λ_{SM}. There are two issues with this proposal:

1. Why might m_u be so small?
2. We can measure m_u (with the help of the lattice). Is this consistent with lattice results?

Accounting for small m_u

Banks, Nir, Seiberg put forward models which, in accounting for quark flavor, gave rise to small or zero m_u.

A simple possibility is suggested by string theory, which often exhibits anomalous discrete symmetries; more precisely, the chiral content of the theory is anomalous, with the anomaly being cancelled by the non-linear transformation of an axion-like field. In the supersymmetric case, this means that one has a modulus field, coupling to the $\bar{u}$ quark as ($\Phi = \phi + ia$)

$$e^{-\Phi} Q H_U \bar{u}. \tag{45}$$

One requires that the exponential be very small, but this is plausible. One can speculate as to whether or not a suitable discrete symmetry structure is typical of underlying theories.

How might $m_u = 0$ be consistent with known facts of hadron physics

Instantons suggestive (Georgi-McArthur). With three light quarks, generate an effective u quark mass (two point function) proportional to $m_d m_s$. Simple dimensional analysis suggests the effect goes as

$$\frac{m_d m_s}{\Lambda} \tag{46}$$

with Λ a suitable QCD scale. This could easily be of order the few MeV expected from current algebra. Kaplan and Manohar expressed this as an ambiguity in current algebra, i.e. they isolated a term and second order in quark masses which could mimic the effects of a u quark mass.

Summary of lattice results for light quark masses

Current results from lattice simulations (summarized by the FLAG working group) are inconsistent with $m_u = 0$.

$$m_u = 2.16\ (9)(7)\mathrm{MeV} \quad m_d = 4.68\ (14)(7)\mathrm{MeV} \tag{47}$$

$$m_s = 93.5(2.5)\mathrm{MeV}$$

Numbers are in $\overline{MS}$ scheme at 2 GeV.

So m_u is many standard deviations from zero. Probably end of story, but some proposals for dedicated tests (Kitano), calibrations (Dine, Draper, Festuccia).

Loop Corrections at Low Energies in the Standard Model

Loop corrections to θ in the Standard Model are highly suppressed. Focussing on divergent corrections, one requires Higgs loops. These involve the Hermitian matrices

$$A = y_d^\dagger y_d; \quad B = y_u^\dagger y_u \tag{48}$$

Contributions to θ are proportional to traces of the form

$$Tr(ABA^2B\dots) \tag{49}$$

one additional matrix factor for each loop.

It is easy to check that the first complex combination involves six matrices, e.g.

$$\mathrm{Tr}(ABA^2B^2) \tag{50}$$

but this and its complex conjugate both appear with the same coefficient. It is necessary to add a $U(1)$ gauge loop (which distinguishes u and d) to have the possibility of a complex traces. [Ellis, Gaillard]

So if θ is small at some scale Λ_{SM}, further corrections are extremely tiny (finite corrections are also very small).

One has the feeling that this might not be such a big problem. The question is: why might $\theta(\Lambda_{SM}$ be small?

Spontaneous CP Violation: The Nelson-Barr mechanism

Invokes spontaneous CP violation to argue “bare θ" is zero. Constructs a mass matrix such that spontaneous CP breaking gives a large CKM angle (as observed, $\delta = 1.2$) with $\arg \det m_q = 0$.

Bare θ is tree level θ (presumes some perturbative approximation). Must insure that $\theta(\Lambda_{SM})$ is small.

Unlike axion, $m_u = 0$ solutions, no obvious low energy consequences.

Attempts to achieve a setup where θ at the scale Λ_{SM} is extremely small.

Such a structure is perhaps made plausible by string theory, where CP is a (gauge) symmetry, necessarily spontaneously broken. At string scale, $\theta = 0$ a well-defined notion. Some features of the required mass matrices appear, e.g., in Calabi-Yau compactifications of the heterotic string.

Simple realization of the NB structure

Complex scalars η_i with complex (CP-violating) vev's.
Additional vectorlike quark with charge $1/3$.

$$\mathcal{L} = \mu \bar{q} q + \lambda_{if} \eta_i \bar{d}_f q + y_{fg} Q_f \bar{d}_g \phi \tag{51}$$

where ϕ is Higgs; y, λ, μ real.

$$M = \begin{pmatrix} \mu & B \\ 0 & m_d \end{pmatrix} \tag{52}$$

$B_f = \lambda_{if} \eta_i$ is complex. M has real determinant.

The structure is reminiscent of an E_6 gauge theory, which has the requisite vector-like quarks and singlets.

Requirements for a successful NB Solution

1. Symmetries: It is important that η_i not couple to $\bar{q}q$, for example. So, e.g., η's complex, subject to a Z_N symmetry.
2. Coincidences of scale: if only one field η, CKM angle vanishes (can make d quark mass matrix real by an overall phase redefinition). Need at least two, and their vev's (times suitable couplings) have to be quite close:

$$\delta_{CKM} \propto \frac{B_{small}}{B_{large}} \tag{53}$$

3. Similarly, μ (which might represent vev of another field) can not be much larger than η_i, and if much smaller the Yukawa's and B's have to have special features.

Constraints on the Overall Scale

Before considering radiative effects, possible higher dimension operators in $\mathcal{L}$ constrain the scales η_i, μ. E.g.

$$\frac{\eta_i^* \eta_j}{M_p} \bar{q} q \tag{54}$$

requires $\frac{|\eta|}{M_p} < 10^{-10}$.

Barr-Nelson With/Without Supersymmetry

Without supersymmetry, highly tuned. Two light scalars and μ (or three light scalars), with masses 10 orders of magnitude below M_p. Far worse than θ.

Even ignoring that, require close coincidence of scales.

Supersymmetry helps. Allows light scalars. Coincidences still required (and more chiral multiplets to achieve desired symmetry breakings – typically at least seven). Some of the high dimension operators better controlled (e.g. if μ, η_i much larger than susy breaking scale, don't have analogs of the $\eta_i^* \eta_j \bar{q} q$ operator).

Loop Corrections in Nelson-Barr: Non-Supersymmetric case

In the non-supersymmetric case, in the simplest model, potential corrections arise at one loop order. Consider, in particular, couplings of the form

$$\lambda_{ij}\eta_i\eta_j|H|^2$$

give rise to one loop contributions.

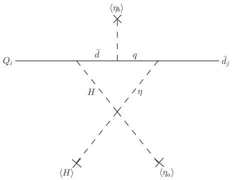


Figure 1: Example threshold correction to Arg det m_d.

If the new couplings are of order one these are six or seven orders of magnitude too large.

In the past these have sometimes been dismissed on the grounds that these couplings contribute to the Higgs mass, but this is just part of the usual fine tuning problem.

Supersymmetry breaking and Nelson-Barr

Many possible phases once allow soft breaking **Note: these effects don't decouple for large susy-breaking scale**. E.g. is susy breaking described by Goldstino superfield, X, superpotential couplings

$$\frac{\mathcal{O}_d}{M_p^{d-2}} X \tag{55}$$

where $\langle \mathcal{O} \rangle$ is complex can lead to large phases in soft breakings. Similarly phases in W. Phases in gaugino masses feed directly into θ.

Loop Corrections in Supersymmetric Nelson-Barr

If tree level phases in soft terms suppressed, loops still pose a problem (Kagan, Leigh, M.D.). Loop corrections to gaugino mass from loops with q, $\bar{q}$, fields. Require, e.g., A terms small or proportional to Yukawas. Gauge mediation (with real F) most plausible solution (A terms small). (Luty, Schmaltz in a slightly different context)

The Peccei-Quinn Symmetry

In a somewhat streamlined language, the Peccei-Quinn proposal was to replace θ by a dynamical field: $\theta \to \frac{a(x)}{f_a}$

It is assumed that $a \to a + \omega f_a$ is a good symmetry of the theory, *violated only by effects of QCD*. Without QCD, θ can take any value.

In QCD *by itself*, the energy is necessarily stationary when

$$\theta_{eff} = \langle \frac{a}{f_a} \rangle = 0. \tag{56}$$

This is simply because *CP* is a good symmetry of QCD if $\theta = 0$, so the vacuum energy (potential) must be an odd function of θ.

One can do better, calculating, again using what we know about chiral symmetry in QCD, the axion potential:

$$V(a) = m_\pi^2 f_\pi^2 \frac{\sqrt{m_u m_d}}{m_u + m_d} \frac{a^2}{2f_a^2} \tag{57}$$

This gives, for the axion mass:

$$m_a = 0.6 \text{ meV} \left(\frac{10^{10} \text{ GeV}}{f_a} \right). \tag{58}$$

[**Exercise**: Derive the expression for $V(a)$. To do this, integrate out the “heavy" fields (the pions) by solving the π_0 equation of motion, and subsituting back in the action. Simplify by working to second order in a, π^0.]

Peccei and Quinn actually constructed a model for this phenomenon, which was a modest extension of the Standard Model with an extra Higgs doublet. They didn't phrase the problem in quite the way I did above, and didn't appreciate that their model had a light, pseudoscalar particle, a. This was recognized by Weinberg and Wilczek, who calculated its mass and the properties of its interactions. It quickly become clear that the original axion idea was not experimentally viable.

The Invisible Axion

But in the more general picture described above, the problems with the axion are easily resolved. The strength of the axion's interactions are proportional to $1/f_a$. This is because of the Peccei-Quinn symmetry. The symmetry requires that axion interactions appear only with derivatives of the axion field; on dimensional grounds, these come with powers of $\frac{\partial_\mu}{f_a}$ (momenta – q^μ/f_a). QCD terms which break the symmetry also come with powers of $1/f_a$. So if f_a is large enough, the axion will be hard to detect (it becomes "harmless" or "invisible").

The scale, f_a, might be associated with some high scale of physics (M_{gut}? M_p? – more later).

Sample couplings

1. Axion to two photons (notation of PDG):

$$\mathcal{L}_{\gamma\gamma} = \frac{1}{4} G_{a\gamma\gamma} \; a \, F\tilde{F} \tag{59}$$

where now F is the *electromagnetic* field strength.

$$G_{a\gamma\gamma} = \frac{\alpha}{2\pi} \left(\frac{E}{N} - \frac{4}{3} \frac{4+z}{1+z} \right) \frac{1+z}{\sqrt{z}} \frac{m_a}{m_\pi f_\pi} \quad z = \frac{m_u}{m_d} \tag{60}$$

$E, \; N$ are the electromagnetic and QCD anomalies of the PQ current.

2. Axion to quarks, leptons:

$$\mathcal{L}_{aff} = \sum_f \frac{C_f}{2f_a} \bar{\psi}_f \gamma^\mu \gamma_5 \psi_f \; \partial_\mu a. \tag{61}$$

The detailed coefficients depend on the model.

Two Benchmark models

DFSZ

Add to the Standard Model an additional Higgs doublet (e.g. as in supersymmetry), i.e. two doublets, H_u, H_d, plus a singlet, ϕ. Impose the Peccei-Quinn symmetry:

$$\phi \to e^{i\alpha}\phi; \;\; H_u \to e^{-i\frac{\alpha}{2}} H_u; \;\; H_d \to e^{-i\frac{\alpha}{2}} H_d \tag{62}$$

Require potential such that H_U, H_D, ϕ have expectation values, where the ϕ vev is very large.

$$\langle\phi\rangle \approx \frac{f_a}{\sqrt{2}} \gg \mathrm{TeV}. \tag{63}$$

This breaks the PQ symmetry spontaneously. (Pseudo-)Goldstone boson:

$$\mathrm{Im}\ \phi = \frac{a}{\sqrt{2}}.$$

a couples to $G\tilde{G}$, $F\tilde{F}$. Also couples to leptons, quarks.

$$\frac{E}{N} = 8/3; \quad C_e = \frac{\cos^2\beta}{3} \quad \tan\beta = \frac{\langle H_u \rangle}{\langle H_d \rangle}. \tag{64}$$

As expected, as f_a becomes large, the axion's interactions with other particles become weaker. Once $f_a >> \mathrm{Tev}$, unobservable in accelerator experiments.

KSVZ Model

Here one has a field, ϕ, and a new quark, q and $\bar{q}$, which will be very heavy. q and $\bar{q}$ are assumed to carry color but to be $SU(2) \times U(1)$ singlets. In two component language, the Peccei-Quinn symmetry is assumed to be

$$\phi \to e^{i\alpha}\phi \quad q \to e^{-i\frac{\alpha}{2}} q \,' \bar{q} \to e^{-i\frac{\alpha}{2}} \bar{q}. \quad \mathcal{L}_{\phi\bar{q}q} = \lambda\phi\bar{q}q \qquad (65)$$

ϕ is assumed to have an expectation value:

$$\langle\phi\rangle = \frac{f_a}{\sqrt{2}}. \qquad (66)$$

The imaginary part of ϕ is the axion:

$$\phi = \frac{1}{\sqrt{2}}\left(f_a + ia\right). \qquad (67)$$

But these are just two of a wealth of possible models, characterized by the coefficients E, N, C_i above. These two, however, are often used as benchmarks to characterize the capabilities of different experimental detection schemes, as well as to illustrate the range of possible astrophysical phenomena.

Physical Processes Associated with Axions

1. Production in accelerators
2. Decay
3. Production in stars
4. Production in strong magnetic fields (ADMX and other experiments

Astrophysical Constraints

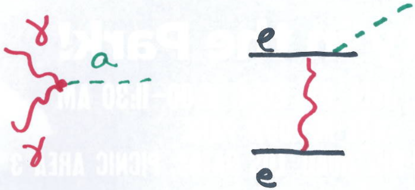
Axion interactions are “semi weak", in the sense that cross sections go as $1/f_a^2$, as opposed to weak interactions which behave as $1/v^4$. So even for large f_a, reaction rates can be comparable to those for neutrinos. This raises a worry about stars, where various processes can produce axions. If interaction rates are large compared to those for neutrinos, excessive amounts of energy will be carried off by axions. More detailed studies in particular astrophysical environments place lower limits on f_a.

Sources of Astrophysical Constraints

Partial list:

1. The sun
2. Red Giants, Globular Clusters
3. SN 1987a
4. White dwarfs

Primakoff process, axion bremstrahlung.



Axion Luminosity

In sun:

$$L_a = G_{a\gamma\gamma}^2 \times 1.85 \times 10^{17} L_\odot \tag{68}$$

so

$$G_{a\gamma\gamma} < 7 \times 10^{-10}. \tag{69}$$

Stronger constraint from globular clusters, $7 \to 1$.