't Hooft and η'ail
Instantons and their applications

Submitted by **Flip Tanedo**

on 4 May 2010 in partial fulfillment of the requirements for the advancement to candidacy for the degree of doctor of philosophy at Cornell University.

Institute for High Energy Phenomenology,
Newman Laboratory of Elementary Particle Physics,
Cornell University, Ithaca, NY 14853, USA

E-mail: pt267@cornell.edu

Prompt

Explain the role of instantons in particle physics

1. As a warm-up explain how instantons describe tunneling amplitudes in quantum mechanics

2. Explain the vacuum structure of gauge theories. In particular explain what the winding number is and how the $|n\rangle$ vacua appear, and show how instantons describe tunneling between these vacua. Explain what the θ vacuum is.

3. As an application, explain the U(1) problem ("η problem") of QCD and how instantons it. Do not simply say that $\mathrm{U}(1)_A$ is anomalous and broken by instantons; explain the Kogut-Susskind mechanism.

4. Explain the emergence of the 't Hooft operator. Explain the relation between anomalies and instantons, focusing on the index theorem, and also show how the 't Hooft operator encodes the breaking of the anomalous symmetries. As an example show how the 't Hooft operator can lead to baryon and lepton number violation in the SM. As another example of the 't Hooft operator consider $\mathcal{N} = 1$ SUSY QCD with$F = N - 1$ flavors (for and SU(N) gauge group). Write down the 't Hooft operator for that theory, and explain how that could actually come from a term in the superpotential, and why $F = N - 1$ is special.

Contents

1 *Une dégustation*: Introduction

Instantons play a rather understated role in standard quantum field theory textbooks, tucked away as an additional topic or mentioned in passing. Despite being notoriously difficult to calculate, instanton effects play a very important role in our conceptual understanding of quantum field theory.

In this examination we provide a pedagogical introduction to the instanton and some of its manifestations in high energy physics. We begin with 'instanton' configurations in quantum mechanics to provide a controlled environment with few 'moving parts' that might distract us from the real physics. After we've used path integrals and imaginary time techniques to remind ourselves of several quantum mechanical facts that we already knew, we will make use of this foundation to understand the basis of Yang-Mills instantons and how tunneling between degenerate vacua can be [surprisingly] manifested in quantum field theory. We will discover that Yang-Mills theory has a surprisingly rich vacuum structure that can perform seemingly miraculous feats in the presence of fermions. In particular, we will see how Yang-Mills instantons (which *a priori* have nothing to do with fermions) can solve an apparent problem in the spectrum of low-lying spectra of hadrons. In doing this we will present a deep relationship between instantons and anomalous U(1) symmetries that can be used for baryon and lepton number violation in the early universe. We will end with an introduction to the moduli space of SUSY QCD where instanton effects end up providing a powerful handle to solve the ADS superpotential for any number of flavors and colors. By the end of this paper we hope that the pun in its title will be clarified to the reader.

There are many things that we will unfortunately not be able to cover, but hopefully this may serve as an introduction to a fascinating subject.

This work is a review of a broad subjects and we've made very little attempts to provide references to original literature. In the appendix we give a brief literature review of the sources that this paper drew most heavily upon. Of all the sections of this paper, the literature review is probably the most useful to other graduate students hoping to learn more about instantons. Nothing in this work represents original work, except any errors. Of course, at the time of compilation this document was perfect and flawless; any errors must have occurred during the printing process.

This examination was prepared over the course of a three week period along with two other similar exam questions. The reader will note that the tone in the document becomes increasingly colloquial and borderline sarcastic as the deadline approached. (The introduction was written last.)

That being said, *Allons-y!*

2 *Apéritif*: Quantum mechanics

Let's start with the simplest instanton configuration one could imagine: tunneling in quantum mechanics, i.e. 0+1 dimensional quantum field theory. This is a story that we are all familiar with from undergraduate courses, but we will highlight the aspects which will translate over to Yang Mills theory in 4D. We will summarize the presentation by Coleman [1]; for details see also [2].

2.1 The semiclassical approximation and path integrals

Exact solutions to quantum systems are intractably difficult to solve. To make progress, we proceed using perturbation theory about the analytically soluble (i.e. quadratic) classical path[1]. This corresponds to an expansion in $\hbar$ (in the language of Feynman diagrams, an expansion in loops).

In quantum mechanics we know how to take the semiclassical limit: the WKB approximation. Here we note that for a constant potential, the solution to the Schrödinger equation is just a plane wave,

$$\Psi(x) = \Psi(0)e^{\pm ipx/\hbar} \qquad\qquad p = \sqrt{2m(E-V)}. \tag{2.1}$$

For nontrivial $V \to V(x)$, we can promote $p \to p(x)$ and proceed to solve for $p(x)$. This plane-wave approximation is valid so long as the state is probing a 'sufficiently flat' region of the potential. This can be quantified by defining the [position-dependent] Compton wavelength of the quantum state

$$\lambda(x) = \frac{2\pi\hbar}{p(x)}. \tag{2.2}$$

Thus we can see that the $\hbar \to 0$ limit indeed allows the quantum state to probe smaller (and hence more slowly varying) regions of the potential.

From here one can proceed as usual from one's favorite introductory quantum mechanics textbook to derive nice results for scattering and even tunneling. Since we're grown ups, let us instead remind ourselves how this comes about in the path integral formalism. Feynman taught us that quantum ampilitudes can be written as a sum over paths

$$\langle x_f|e^{iHt/\hbar}|x_i\rangle = N\int [dx]e^{-iS/\hbar}. \tag{2.3}$$

The semiclassical limit $\hbar \to 0$ causes the exponential on the right-hand side to oscillate quickly so that nearby paths tend cancel one another. The [parametrically] dominant contribution to the path integral then comes from the path of stationary phase (steepest descent). This, of course, comes from the extremum of the action and corresponds to the classical path, x_{cl}. The usual game, then, is to expand about this path

$$x(t) = x_{\text{cl}}(t) + \sum_n c_n x_n(t) \tag{2.4}$$

using a convenient basis of functions $x_n(t)$ chosen so that the resulting functional determinant can be calculated easily. *Ho hum!* This is all familiar material since we're all grown up and already know all about fancy things like path integrals.

One thing that might cause us to pause, however, is to ask how this can possibly give us quantum tunneling when—by definition—for such processes there exists no classical path to perturb

[1]The identification of the quadratic part of the Lagrangian as classical is most easily seen in $\lambda\phi^4$ theory by scaling $\phi \to \phi' = \lambda^{1/4}\phi$ and noting that the partition function contains an exponential of $\mathcal{L}'/\hbar = (1/\lambda\hbar)[\frac{1}{2}(\partial\phi)^2 + \cdots]$. The semiclassical limit corresponds to small $(\lambda\hbar)$.

about. Conceptually this is a bit of a hum-dinger, but a hint can already be seen from (2.1): for $E < V$ the momentum becomes imaginary and we get the expected exponential behavior. Going back to our high-brow path integrals, we shall now see that we can get to this behavior by working in the imaginary time formalism via our old friend, the Wick rotation,

$$t = i\tau. \tag{2.5}$$

The validity of the Wick rotation may seem odd, but we shall simply treat this as a change of variables[2]. What does this buy us? Our amplitudes now take the form

$$\langle x_f|e^{-H\tau/\hbar}|x_i\rangle = N\int[dx]e^{-\frac{1}{\hbar}\int^{\tau}\mathcal{L}_{\mathrm{E}}d\tau'} \tag{2.6}$$

where

$$\mathcal{L}_{\mathrm{E}} = \frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 + V(x). \tag{2.7}$$

We see that the potential has swapped signs relative to the kinetic term. This is easy to see from the equation of motion, $m\ddot{x} = -V'(x)$, where the left-hand side picks up an overall sign when $t \to \tau$. On the surface this seems like a completely trivial change (it is), but the point is that the minus sign flips the potential barrier upside down allows us to find a classical path between the two extrema. The imaginary time formalism provides a classical path about which we can sensibly make a semiclassical approximation for tunneling processes.

Let us make two remarks:

- The [imaginary] time evolution operator $\exp(-H\tau/\hbar)$ is not unitary, but Hermitian.
- An interesting consequence of this is that large time evolution leads to a projection to the ground state (if the state has any overlap with $|0\rangle$):

$$\lim_{\tau\to\infty}\langle x\left|e^{-H\tau/\hbar}\right|x'\rangle = \lim_{\tau\to\infty}\langle x|n\rangle\langle n|x'\rangle e^{-E_n\tau/\hbar} \tag{2.8}$$

$$= \langle x|0\rangle\langle 0|x'\rangle e^{-E_0\tau/\hbar} \tag{2.9}$$

$$= \psi_0(x)\psi_0^*(x')e^{-E_0\tau/\hbar}. \tag{2.10}$$

2.2 The harmonic oscillator

Before discussing any tunneling phenomena, let's review the harmonic oscillator in the imaginary time formalism. In all cases to follow we will consider events between an initial time $-T/2$ and a final time $T/2$ with $T \to \infty$. As before, we shall expand about the classical path

$$x(\tau) = x_{\mathrm{cl}}(\tau) + \sum_n c_n x_n(\tau), \tag{2.11}$$

[2] Analyticity is a rather deep idea in physics which the author is still trying to appreciate with the proper reverence. For example, non-analyticity in the effective potential signals the appearance of massless modes. More mundanely, in the Kramers-Kronig relation analyticity is related to causality. In the present case, the important feature is that the Hamilton-Jacobi equations still hold upon complexification. For more about this rather deep connection, see e.g. [3].

where we choose our basis functions x_n to satisfy the appropriate boundary conditions ($x_n(\pm T/2) = 0$) and to be eigenfunctions of $\delta^2 S/\delta x_{\rm cl}^2$,

$$-\frac{d^2 x_n}{x\tau^2} + V''(x_{\rm cl})x_n = \lambda_n x_n. \tag{2.12}$$

The path integral measure is converted to $[dx] \to \prod_n (2\pi\hbar) dc_n$, where we use Coleman's normalization [1]. In the semiclassical limit where $\hbar \to 0$, stationary phase tells us that the amplitude is dominated by

$$\langle x_f | e^{-H\tau/\hbar} | x_i \rangle = N e^{-S(x_{\rm cl})/\hbar} \left[\det\left(-\partial_\tau^2 + V''(x_{\rm cl})\right)\right]^{-1/2} \tag{2.13}$$

up to higher order terms in $\hbar$. By our choice of basis functions the determinant over $\delta^2 S/\delta x_{\rm cl}^2$ can be written as $\prod_n \lambda^{-1/2}$. With suave arithmetic maneuvering one can show that the normalization and the determinant simplify to

$$\left[\det\left(-\partial_\tau^2 + V''(x_{\rm cl})\right)\right]^{-1/2} = \left(\frac{\omega}{\pi\hbar}\right)^{1/2} e^{-\omega T/2}, \tag{2.14}$$

where $\omega \equiv V''(x_{\rm cl})$. Note that this agrees with our earlier remark that a long time evolution projects out the ground state. The derivation of this expression is not particularly enlightening, but (2.14) will be a useful result for future reference. A proper derivation can be found in [2].

Proof. **Derivation of (2.14).** We shall follow the steps in [2]. The boundary conditions on our basis functions allow us to solve for their eigenvalues,

$$\lambda_n = \frac{\pi n^2}{T^2} + \omega^2. \tag{2.15}$$

Let us then massage the factors in (2.14) as follows

$$\left[\det\left(-\partial_\tau^2 + V''(x_{\rm cl})\right)\right]^{-1/2} = N \prod_n \lambda_n^{-1/2} \tag{2.16}$$

$$= N \prod_n \left(\frac{\pi^2 n^2}{T^2}\right)^{-1/2} \prod_n \left(1 + \frac{\omega^2 T^2}{\pi^2 n^2}\right)^{1/2}. \tag{2.17}$$

We have simply pulled out the first factor from the expression for each λ_n. This term should look rather familiar. It is the only term to survive the limit $V \to V_0 =$ constant , i.e. it gives the 'classical' contribution for a plane wave solution. We have

$$N \prod_n \left(\frac{\pi^2 n^2}{T^2}\right)^{-1/2} = \int \frac{dp}{2\pi} e^{-p^2 T/2} = \frac{1}{\sqrt{2\pi T}}. \tag{2.18}$$

Meanwhile, for the other factor we can invoke an identity for the hyperbolic sine,

$$\sinh(\pi y) = \pi y \prod_n \left(1 + \frac{y^2}{n^2}\right). \tag{2.19}$$

This allows us to write

$$\left[\det\left(-\partial_\tau^2 + V''(x_{\rm cl})\right)\right]^{-1/2} = \frac{1}{\sqrt{2\pi T}}\left(\frac{\sinh(\omega T)}{\omega T}\right)^{-1/2} \tag{2.20}$$

$$= \sqrt{\frac{\omega}{\pi}}\left(2\sinh(\omega T)\right)^{-1/2} \tag{2.21}$$

$$= \sqrt{\frac{\omega}{\pi}} e^{-\omega T/2}\left(1 + \frac{1}{2}e^{-2\omega T} + \cdots\right). \tag{2.22}$$

□

2.3 The double well

Now that we're warmed up, let's move on to a quantum mechanical system with actual tunneling. As we discussed, the imaginary time formalism flips over the potential so that there now exists a classical path between the two extrema. This is shown heuristically in Fig. 1. The classical equation of motion tells us that if we start at $x = -a$, there is a conserved energy

$$E = 0 = \frac{1}{2}\left(\partial_\tau x\right)^2 - V(x). \tag{2.23}$$

From this we get $\partial_\tau x = \sqrt{2V(x)}$, which we can integrate.

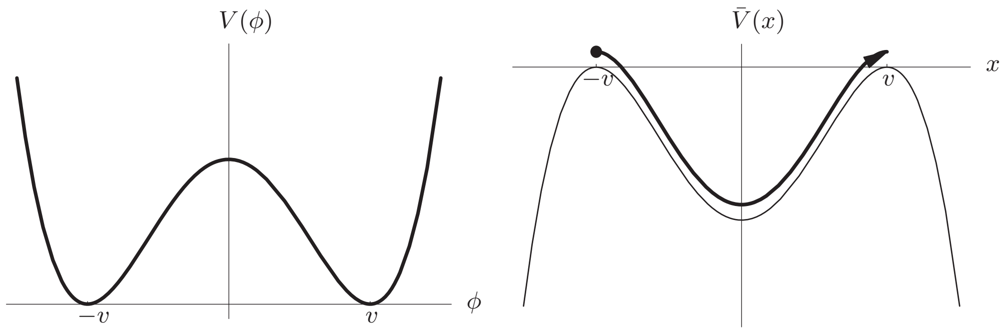

Figure 1: The double well potential and its flipped over euclidean version, from [4]. We use slightly different notation: the extrema will be labelled $x = \pm a$.

We start by writing out the action associated with this 'classical' tunneling path,

$$S_0 = \int d\tau \left[\frac{1}{2}(\partial_\tau)^2 + V(x)\right] = \int dt \left(\frac{dx}{dt}\right)^2 = \int_{-a}^{a} dx\sqrt{2V}. \tag{2.24}$$

Taylor expanding $\sqrt{2V}$ about $x = a$ at late times tells us that for $\tau \gg 1$

$$\partial_\tau x \approx \omega(a - x) \Rightarrow (a - x) \propto e^{-\omega\tau}. \tag{2.25}$$

As before, we've written $\omega = V''(a)$. This tells us that the instanton solution is a kink of some characteristic width $1/\omega$; these topological configurations are localized in time. This is the origin of the term **instanton**, or (as Polyakov suggested), "pseudo-particle."

One can solve for the structure of the instanton kink solution that interpolates between the two vacua. The answer is, as we would expect, a hyperbolic tangent. Instead of dwelling on this, let us move on to consider the tunneling amplitude. The usual formula with $S(x_{\text{cl}}) = S_0$ gives

$$\langle a|e^{-H\tau/\hbar}|-a\rangle = Ne^{-S_0/\hbar}\left[\det\left(-\partial_\tau^2 + V''(x_{\text{cl}})\right)\right]^{-1/2}. \tag{2.26}$$

In the case of the harmonic oscillator (single well) we already solved for the normalized determinant and found (2.14). Since the two minima in the double well locally behave like single wells and since the quantum state spends *most* of its time very close to one or the other well, our physics intuition tells us that (2.14) should hold up to some corrections to account for the instanton. Let us parameterize this correction as an overall factor K,

$$N\left[\det\left(-\partial_\tau^2 + V''(x_{\text{cl}})\right)\right]^{-1/2} = \left(\frac{\omega}{\pi\hbar}\right)^{-1/2} e^{-\omega T/2}K. \tag{2.27}$$

Thus we find that for a single instanton background,

$$\langle a|e^{-H\tau/\hbar}|-a\rangle = \left(\frac{\omega}{\pi\hbar}\right)^{-1/2} e^{-\omega T/2}Ke^{-S_0/\hbar}. \tag{2.28}$$

2.3.1 The zero mode

This isn't the whole story. For large times (e.g. compared to the characteristic instanton lifetime $1/\omega$) we should properly account for the time translation invariance; i.e. we must include the effects of instantons occurring at any intermediate time. We can pitch this in fancier language. For example, we can say that the instanton has a **collective coordinate** which acts as a parameter to give different instanton solutions. In other words, the instanton has a **moduli space**.

More physically, we remark that the mode x_0 associated with this time translation can be thought of as a **Goldstone boson** for the spontaneously broken τ-symmetry. As one would expect, such a Goldstone mode has zero eigenvalue: $\lambda_0 = 0$. We can write out the form of this mode explicitly:

$$x(\tau) = x_{\text{cl}}(\tau + d\tau) = x_{\text{cl}}(\tau) + x_{\text{cl}}(\tau + d\tau) - x_{\text{cl}}(\tau) = x_{\text{cl}}(\tau) + \frac{dx_{\text{cl}}}{d\tau}d\tau + \cdots, \tag{2.29}$$

so that comparing to (2.11), we have

$$x_0 = S_0^{-1/2}\frac{dx_{\text{cl}}}{d\tau}, \tag{2.30}$$

where the normalization comes from our normalization of the classical path in (2.24) and the requirement that the modes x_n must be orthonormal.

We've made a bed of fancy words and now we have to lie in it. In particular, now that we've thrown around the idea of a **zero mode**, we need to stop and go all the way back to our evaluation of the functional integral. When we evaluated the functional integral into a determinant, we assumed that all of the eigenvalues λ_n were positive definite. This made all of our Gaussian integrals sensible. Now we have a zero eigenvalue, we have an apparent conundrum: $\exp(-c_0^2\lambda_0)dc_0 = dc_0$, so that we no longer have a Gaussian integral! This integral is formally infinite and we worry that our state is not normalizable.

Fortunately, we have nothing to fear. This apparently-divergent integral is simply corresponds to the integration over the instanton time that we mentioned above. In fact, in this light it is an *expected* 'divergence' for we need some kind of extensive dependence on the large time T.

To be more rigorous, we would like to convert the dc_0 integral into a $d\tau_c$ where τ_c is the instanton center. This is easy to work out since we know that the change in the instanton configuration $x(\tau) = x_{\text{cl}}(\tau) + \sum_n c_n x_n$ induced by a small change in τ_c is

$$dx = \frac{dx_{\text{cl}}}{d\tau}\, d\tau_c. \tag{2.31}$$

Meanwhile, the change in the instanton configuration from a change in c_0 is

$$dx = x_0\, dc_0. \tag{2.32}$$

Combining these results we find that

$$dc_0 = \sqrt{S_0}\, d\tau_c. \tag{2.33}$$

This wasn't particularly surprising, but we want to keep track of the constant of proportionality ($\sqrt{S_0}$) since it will feed into our determination of the factor K. This digression is prescient in another way: it is an example of how an instanton couples to zero modes of a system, and so is a crude prototype for the **'t Hooft operator** that we will explore in QCD.

2.3.2 Determining K

While the current treatise focuses primarily on the qualitative features of what instantons are and what they can do, it is nice to flesh out quantitative results when they don't require much work[3]. For our quantum mechanical example, it is nice to explicitly work out the K factor in (2.28). From the zero mode discussion we found that

$$\frac{1}{\sqrt{2\pi\hbar}}\, dc_0 = \sqrt{\frac{S_0}{2\pi\hbar}}\, d\tau_c. \tag{2.34}$$

Thus when we take into account the integration over the instanton position, our tunneling amplitude takes the form

$$\langle a|e^{-HT/\hbar}|-a\rangle = NT\sqrt{\frac{S_0}{2\pi\hbar}} e^{-S_0/\hbar}\left[\det\left(-\partial_\tau^2 + V''(x_{\text{cl}})\right)\right]^{-1/2}. \tag{2.35}$$

Comparing to (2.28) we find that

$$K = \sqrt{\frac{S_0}{2\pi\hbar}}\left|\frac{\det(-\partial_\tau^2+\omega^2)}{\text{Det}(-\partial_\tau^2 + V''(x_{\text{cl}}))}\right|, \tag{2.36}$$

[3] Unfortunately, most of the 'honest' calculations in this field require a lot of work.

where we've introduced the notation Det to mean determinant *without* any zero modes. (Most review literature refers to this as det', but the author thinks this notation is silly because it is suggestive of some kind of derivative.)

2.3.3 Multi-instanton effects

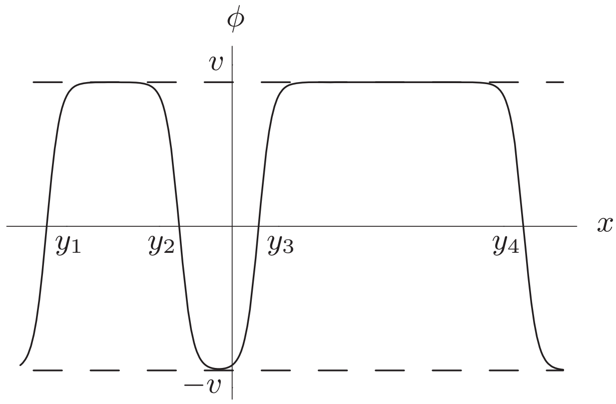

Figure 2: An example of a multi-instanton background from [4]. Note that we use different notation. The author is too pressed for time to draw his own diagrams properly.

We're not yet done with the double well. If we think about it a bit longer, we'll note that the one instanton solution is not the only possible classical background to perturb about. One could, in fact, have a chain of instantons and 'anti-instantons' tunneling back and forth between the vacua, as depicted in Fig. 2. The only restriction is that instantons from $|-a\rangle \to |a\rangle$ can only be followed by anti-instantons from $|a\rangle \to |-a\rangle$ and vice versa. Our previous logic leading up to the K factor should tell us that these n-instanton solutions should take the same form but with $K \to K^n$:

$$N\left[\det(-\partial_\tau^2 + V''(x_{\rm cl}))\right]^{-1/2} = \left(\frac{\omega}{\pi\hbar}\right)^{-1/2} e^{-\omega T/2} K^n e^{-S_0/\hbar}. \tag{2.37}$$

One should again integrate over these instanton positions (corresponding to their zero modes, as discussed above). Note that the alternating instanton/anti-instanton ordering gives us a slightly more non-trivial integral,

$$\int_{T/2}^{T/2} d\tau_1 \int_{-T/2}^{\tau_1} d\tau_2 \cdots \int_{-T/2}^{\tau_{n-1}} d\tau_n = \frac{T^n}{n!}. \tag{2.38}$$

This integral should look rather familiar from one's first introduction to perturbation theory and canonical quantization in quantum field theory.

The 'full' classical background for the semiclassical approximation must take into account these multi-instanton solutions. We thus find

$$\langle \pm a|e^{-H\tau/\hbar}|-a\rangle = \sqrt{\frac{\omega}{\pi\hbar}}e^{-\omega T/2}\sum_{n\text{ odd/even}}\frac{\left(Ke^{-S_0/\hbar}T\right)^n}{n!} \tag{2.39}$$

$$= \sqrt{\frac{\omega}{\pi\hbar}}e^{-\omega T/2}\frac{1}{2}\left[e^{Ke^{-S_0/\hbar}T} \mp e^{-Ke^{-S_0/\hbar}T}\right] \tag{2.40}$$

up to leading order in $\hbar$. Recalling that the time evolution operator has eigenstates of definite energy, e.g. (2.8), we find that the energy eigenstates and their eigenvalues are

$$|\pm\rangle \equiv |a\rangle \pm |-a\rangle \qquad\qquad E_\pm = \frac{1}{2}\hbar\omega \pm \hbar K e^{-S_0/\hbar}. \tag{2.41}$$

This is, of course, exactly as we would have expected since the action obeys a $\mathbb{Z}_2$ symmetry under which $x \to -x$. In such a case we know that the ground state of the system is degenerate with a small breaking coming from tunneling effects (exactly what we've re-derived). The energy eigenstates of the system are also eigenstates of the parity operator (since this commutes with the Hamiltonian) and the lowest states correspond to the symmetric and antisymmetric combinations. We note that the factor of $\exp(-S_0/\hbar)$ makes it clear that even though this term is small, it is clearly a *non-perturbative* effect that one would not have found doing naïve perturbation theory.

2.4 The periodic potential

There's one more obvious tunneling generalization in quantum mechanics. Since we've already made the leap form a single well harmonic oscillator to a double well potential, it's trivial to go to a triple or n-tuple well potential. The strategy is now completely analogous, though finding closed form solutions for the appropriate instanton sums become non-trivial; e.g. for a triple well one may have up to two instantons in a row but no more. One way to avoid this is to identify the sides of a well, so that we are led to the periodic (cosine) potential.

Remark: It should be 'obvious' that the functional form near the minima of each of these potentials is arbitrarily close to the harmonic oscillator for long times T so that our strategy of duplicating the single well result up to a K factor is justified in the semiclassical approximation.

For the periodic potential there is no constraint on how many instantons (n) or anti-instantons ($\bar{n}$) one might have nor is there any constraint on their ordering except that they must sum to give the appropriate shift in vacua. For example, consider the tunneling process $|j\rangle \to |k\rangle$. The appropriate generalization of our above techniques is

$$\langle k|e^{-HT/\hbar}|j\rangle = \sqrt{\frac{\omega}{\pi\hbar}}e^{-\omega T/2}\sum_{n=0}\sum_{\bar{n}=0}\frac{1}{n!\bar{n}!}\left(Ke^{-S_0/\hbar}T\right)^{n+\bar{n}}\delta_{(n-\bar{n}),(k-j)}. \tag{2.42}$$

We may now use one of our usual tricks and go to a Fourier series representation,

$$\delta_{ab} = \int_0^{2\pi}\frac{d\theta}{2\pi}\,e^{i\theta(a-b)}. \tag{2.43}$$

This allows us to simplify our amplitude to the form

$$\langle k|e^{-HT/\hbar}|j\rangle = \sqrt{\frac{\omega}{\pi\hbar}}e^{-\omega T/2}\int_0^{2\pi}\frac{d\theta}{2\pi}\,e^{i(j-k)\theta}\exp\left(2KTe^{-S_0/\hbar}\cos\theta\right). \tag{2.44}$$

Just as the double well potential contained a discrete permutation symmetry that forced the energy eigenstates to also be permutation eigenstates (whose energy is only split by tunneling effects), the periodic potential also contains a translation symmetry that forces eigenstates to be shift-invariant. This should again sound very familiar from toy models of solid state systems: the energy eigenstates are **Bloch waves**,

$$|\theta\rangle = \left(\frac{\omega}{\pi\hbar}\right)^{1/4}\frac{1}{\sqrt{2\pi}}\sum_n e^{-in\theta}|n\rangle. \tag{2.45}$$

The $|\theta\rangle$ states are eigenstates of a shift operator, e.g. a shift to the next well to the right, T_1:

$$T_1|\theta\rangle = e^{i\theta}|\theta\rangle. \tag{2.46}$$

The energy eigenvalues for the Bloch waves are

$$E(\theta) = \frac{1}{2}\hbar\omega + 2\hbar K e^{-S_0/\hbar}\cos\theta. \tag{2.47}$$

This example will turn out to be very handy when we discuss the θ-vacua of Yang-Mills theories.

2.5 The bounce

If you are doing everything well, you are not doing enough.
– Howard Georgi, personal motto [5]

One topic that is egregiously omitted in this document is the treatment of 'the bounce' and tunneling from metastable vacua. Such field theoretic calculations have been important in early universe cosmology and, more recently, the long-term stability of vacua in metastable SUSY-breaking models. This type of tunneling is also described very well by first considering a quantum mechanical archetype which demonstrates many subtle aspects of the semiclassical approximation. This is described in Section 2.4 of Coleman's lectures [1].

2.6 Polemics: tunneling in QM versus QFT

2.6.1 Summary of QM

Before wiping our hands of quantum mechanics and graduating to quantum field theory (straight to gauge theory, no less), let's pause to make a very important philosophical point about the transition from 0+1 dimensions to nontrivial spacetimes (i.e. QM $\rightarrow$ QFT). It is not a stretch to say that tunneling is one of the key results in quantum mechanics. The idea that a quantum state can pass through a classically-impenetrable barrier is the foundation for all of the manipulations we did above[4]. This is also related to the idea that energy eigenstates ought to also be eigenstates of any discrete symmetry relating theory's vacua, hence the appearance of symmetric and antisymmetric $|\pm\rangle$ states in the double well or the Bloch waves $|\theta\rangle$ in the periodic potential. The *cost* of barrier penetration was evident from our WKB formula (2.1): the oscillating plane wave picks picks up a factor of i in its argument from $\sqrt{E-V}$ and then becomes exponentially suppressed. The bigger the energy difference the more strongly suppressed the tunneling amplitude. This is all as we have grown to know and love: tunneling is one of the highlight calculations in any quantum mechanics course.

2.6.2 Zero (and one) spatial dimensions is special

In quantum field theory of *any* nonzero space dimension, however, we **never** talk about a physical field tunneling. Never (well, almost—certainly never in an introductory course). The reason is clear: quantum field theory is an infinite number of copies of quantum mechanics: there is a coupled QM oscillator at each point in spacetime so that any discussion about the vacuum is not a statement about a Hilbert space, but rather a *Fock* space. Instead of energy, QFT deals with energy *densities* that must be multiplied by the (infinite) volume of spacetime. More concretely: in order to tunnel from one vacuum to another, each of an *infinite number* of QM oscillators must tunnel. This picks up an infinite number of $e^{-\Delta E}$ suppression factors (where ΔE is the characteristic energy difference) which leads to *zero* tunneling probability between Fock space vacua. Thus in QFT there is never any tunneling phenomena between degenerate vacua with a potential barrier. The double well potential in 1+1 dimensional field theory leads to spontaneous symmetry breaking (as do its more famous 'Mexican hat' generalizations in higher dimensions), rather than any parity-symmetric vacuum states. (To be a bit more honest, we will remark below that the case of a single spatial dimension is also 'special' due to kink solutions and should be lumped together with the 0+1 dimensional case[5].)

From this point of view one ought to say that we've so far gone over a very nice review of undergraduate quantum mechanics, but the document should end right here. It's not clear why there should be anything else to be said about 'instantons'—certainly not in field theory where there is no tunneling phenomena to be considered. Thus you should be wondering why there are any more pages to this document at all.

2.6.3 Can we be clever?

The argument above is certainly heuristic. One could ask if we can be clever enough to find a loophole. A good first attempt is to imagine a situation in field theory with metastable vacua tunneling to a ground state via bubble nucleation. Here *finite* volumes of space tunnel and the difference in energy between the true ground state and the metastable state ($\delta E \sim$ volume) versus the surface tension coming from the energy barrier ($\sim$ surface area) can lead to either vacuum decay or bubble collapse depending on the characteristic size of the quantum fluctuation. Such tunneling events can indeed be calculated using fancy instanton methods (though they are unfortunately

[4]Note, for example, that a double well with an infinite finite-width barrier separating the minima will exactly have two degenerate spectra.

[5]We thank Zohar Komargodski for bringing this to our attention.

outside the scope of the present document). These cases, however, avoid the philosophical issue above because one transitions from a metastable vacuum to an energy-favorable vacuum. It should be clear that one can *never* have this sort of bubble nucleation for physical fields between *degenerate* vacua.

We can be a little more clever and consider the Sine-Gordon kink solution. We know that a scalar field in 1+1 dimensions and a double well potential can have kink solutions where the field only has a nonzero vacuum expectation value over a localized position in space. One typically looks for a time-independent solution from which one can reconstruct time-dependence from Lorentz invariance. One could then imagine Wick rotating to try to construct such a kink in the [imaginary] time direction rather than the space direction. Here, however, one still runs into the problem of a vanishing tunneling amplitude because an infinite number of QM oscillators must undergo barrier penetration. Fancier attempts involving just scalar fields will similarly fail for more general grounds: **Derrick's theorem** tells us that scalar fields cannot have solitonic solutions in dimension higher than one. (A short discussion of Derrick's theorem can be found in, e.g. [6].)

We're on the right track. If one wanted to naïvely generalize the Sine-Gordon kink into a vortex for a two dimensional scalar, it is a well-known result that Derrick's theorem manifests itself as a divergence in the energy of the static configuration. For the case of space-like solitons, however, we already know how to evade Derrick's theorem: we introduce gauge fields. Indeed, the vortex solution is given by the winding of a U(1) gauge field. Now we have a handle for how to proceed into QFT.

2.6.4 Gauge redundancy

What is so special about gauge theory that allows us to create solitons? In quantum mechanics where we spoke about the tunneling of a 'physical' state to another 'physical' state and we argued that in QFT it is impossible to pull an infinitely (spatially) extended field over a potential barrier of finite energy density. Gauge theory provides the loophole we wanted because it gives us lots of manifestly *unphysical* degrees of freedom to twist and wrap about. We will see that topological winding about gauge degrees of freedom will lead to classes of distinct gauge vacua. Surprisingly, tunneling can occur between these vacua (since such tunneling doesn't require a physical field being pulled over an energy potential) and this leads to the construction of Bloch waves (θ-vacua) and rather remarkable physical effects.

It is worth spouting some rhetoric about gauge symmetry that amounts to doctrine rather than physics. Gauge symmetry can be thought of in two ways:

1. A gauge symmetry is what one obtains by taking a global ('normal') symmetry and promote it to a local symmetry.

2. A gauge symmetry is a *redundancy* in the way one describes a physical system.

Both doctrines are compatible, but the latter point of view is particularly handy for the philosophical dilemma at hand. Physical states are defined modulo gauge orbits. In other words, gauge transformations form an equivalence class of physical states. The extra degrees of freedom afforded by this gauge redundancy is certainly convenient, but the actual *physical* system is described by modding out the gauge degree of freedom (fixing a gauge).

The Yang-Mills theories that we get when introducing a gauge redundancy[6] can have additional structure due to the gauge symmetry. In particular, we will find that this structure lends itself to multiple vacua that are gauge equivalent but distinguished topologically from one another. One will not be able to continuously deform one topological vacuum to another without pushing the *physical* field over a potential which, as we discussed, is forbidden in QFT. Further, each of these vacua appear to spontaneously break gauge invariance. The 'magic' now is that because gauge degrees of freedom are *redundant* (whether or not they are continuously connected along the vacuum manifold), one is free to construct the analog of our Bloch wave states *in gauge space*: i.e. we can construct gauge-invariant linear combinations of the topologically distinct vacua. These are called the θ vacua and will be the main topic in this paper.

3 *Entrée*: Vacua of gauge theories

In anticipation of non-trivial vacuum structure and tunneling phenomena, we now study Yang-Mills theory in Euclidean spacetime. We'll be a little bit loose with our conventions—we may miss a sign here or there—but the underlying physics will be transparent. There will be some relatively fancy ideas tossed around, but the physical intuition follows precisely the simple quantum mechanical examples above. This is why we invested so much into our quantum mechanical treatment, it gives us a little bit of wiggle room to play it fast and loose now that there are many more moving parts.

First we'll go over the relevant physics to get to the point. Then we'll close with a section that gives just the *slightest* flavor for the mathematical elegance that's running 'under the hood.'

3.1 Euclidean Yang-Mills

We would now like to consider classical solutions to the Euclidean equation of motion. We define our gauge field and generator normalization via

$$A_\mu = igA^a_\mu t^a \qquad\qquad \mathrm{tr}\left(t^a t^b\right) = \frac{1}{2}\delta^{ab}. \tag{3.1}$$

The Euclidean action takes the form

$$S_E = \frac{1}{4g^2}\int d^4x\, F_{\mu\nu}F^{\mu\nu}, \tag{3.2}$$

where we remind ourselves that we're working with a Euclidean metric so that raised and lowered indices are equivalent. The classical equation of motion is

$$D_\mu F^{\mu\nu} = 0. \tag{3.3}$$

[6]One might introduce such a gauge redundancy to describe vector particles in a handy way, or—more technically—to identify (via the R_ξ gauges) the extra degree of freedom that in a massive vector that disappears for a massless vector.