

Precision decay rate calculations in quantum field theoryAnders Andreassen,^{*} David Farhi,[†] William Frost,[‡] and Matthew D. Schwartz[§]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 6 July 2016; published 13 April 2017)

Tunneling in quantum field theory is worth understanding properly, not least because it controls the long-term fate of our Universe. There are, however, a number of features of tunneling rate calculations which lack a desirable transparency, such as the necessity of analytic continuation, the appropriateness of using an effective instead of classical potential, and the sensitivity to short-distance physics. This paper attempts to review in pedagogical detail the physical origin of tunneling and its connection to the path integral. Both the traditional potential-deformation method and a recent, more direct, propagator-based method are discussed. Some new insights from using approximate semiclassical solutions are presented. In addition, we explore the sensitivity of the lifetime of our Universe to short-distance physics, such as quantum gravity, emphasizing a number of important subtleties.

DOI: [10.1103/PhysRevD.95.085011](https://doi.org/10.1103/PhysRevD.95.085011)**I. INTRODUCTION**

Whether the long-term future of the Universe is controlled by the slow freezing and rarefaction of cosmic acceleration or the sudden formation and growth of negative-energy bubbles is a question of visceral appeal even to nonscientists. There are also practical aspects (from a particle-physics point of view) to vacuum stability, such as eliminating models of new physics or motivating new colliders to measure the top mass. Thus, it would be good to know whether the question of stability can even be answered confidently assuming no new physics. This paper provides a survey of some impediments to establishing that confidence. We provide a new perspective on some old methods, such as the connection between the path integral and tunneling, and bring clarity to some recent debates, such as the UV sensitivity of the Universe's lifetime.

One challenge to computing the rate for tunneling out of our metastable vacuum is establishing a systematically improvable framework for computing this rate in the first place. In quantum mechanics, in the absence of any approximate methods, the decay of a given initial wave function can always be calculated by numerically solving Schrödinger's equation. In quantum field theory, one does not have this crutch: Not only is Schrödinger's equation infinite dimensional, but the wave functional (the field-theory analog of the wave function) inspires little physical intuition.

The first few sections of this paper are devoted to reviewing how decay rates are defined, the relevant time scales, and the derivation of various formulas used to

compute them. The traditional method, pioneered by Coleman and Callan [1] (see also [2]), focuses on computing the imaginary part of the matrix element $\langle a|e^{-HT}|a\rangle$, where in quantum mechanics, $|a\rangle$ is a position eigenstate. Because this matrix element is real, one cannot simply take its imaginary part. Rather, one must analytically continue the potential so that the false vacuum is stable, compute the matrix element, then analytically continue back. There are some excellent reviews of this potential-deformation method [3–7]. The method seems to give the right answer, in cases where it can be checked. Nevertheless, some elements of its derivation seem to us in need of further clarification. For example, for physical potentials, which are bounded from below, analytic continuation gives the wrong answer. Instead, the steepest-descent contour passing through the saddle point associated with the false vacuum plays an essential role. We provide our own perspective on this method, which we hope the reader will find illuminating.

Having digested the Callan-Coleman potential-deformation approach, one suspects that there should somehow be a more direct way to connect tunneling rates to the path integral. Such a connection was presented in [8] and is expounded on in Sec. IV. The method introduced in [8] is based on a direct computation of the probability for a particle to propagate through a barrier. It has the advantage of maintaining a closer connection between the underlying physical assumptions, such as the hierarchy of time scales required for the tunneling rate to be well defined, than the potential-deformation method. A summary comparison of the potential deformation and the direct methods is given in Sec. IV D.

To compute a decay rate in the saddle-point approximation, one must find bounces: solutions to the Euclidean equations of motion with some particular boundary conditions. These bounces are functions $\phi(\vec{x}, \tau)$, where Euclidean time τ parametrizes a path through field space.

^{*} anders@physics.harvard.edu[†] farhi@physics.harvard.edu[‡] wfrost@physics.harvard.edu[§] schwartz@physics.harvard.edu

This path is analogous to the most-probable path that a particle passes through a barrier in the WKB approximation [9]. In Sec. VB, we discuss how to think about the functional $U[\phi]$ so that it provides a close analogy with the potential energy barrier in quantum mechanics. Finding exact bounces can be challenging, even numerically. Fortunately, many features of the exact bounces are well described by approximate bounces which can be studied analytically, as we discuss in Sec. VC.

In quantum field theory, metastability can arise from radiative corrections. The famous example of this is the Coleman-Weinberg model, where a stable potential $V = \lambda\phi^4$ turns over and runs negative due to photon loops in scalar QED [10]. It is natural to anticipate that one should therefore use the effective potential, which demonstrates the instability, to compute the tunneling rate. We argue that this is not correct. First of all, corrections to the effective action which vanish for constant fields may contribute equally to a rate as potential terms. Second, using the effective potential double counts the radiative corrections: Particle loops contribute both to V_{eff} and again to the rate. Although these observations are not deep, it is not uncommon to see V_{eff} used as a classical potential to find bounce solutions. The appropriate use of effective actions is discussed in Sec. VI.

Finally, we reduce this to the Standard Model. The lifetime of our Universe has been intensively studied since the tunneling calculations in quantum field theory were first understood [11–17]. Even if we assume the Standard Model is valid up to the Planck scale, one must be sure that quantum gravity cannot invalidate the perturbative decay rate calculation. Current precision measurements and calculations imply that, in the absence of new physics, our Universe will decay through the formation of ultra-tiny bubbles, with radii $R \sim (10^{17} \text{ GeV})^{-1} \sim 10^{-31} \text{ cm}$. This bubble size is essentially determined by the scale where the β function for the Higgs quartic vanishes (we provide a new derivation of this result in Sec. VC). Although 10^{17} GeV is close to the Planck scale, it has been argued that it is far enough below M_{Pl} that quantum gravitational effects on V_{eff} can be ignored [15–18]. It has also been argued that these effects cannot be ignored since the bubble takes trans-Planckian field values at its center [19–22]. The latter conclusion has been verified by other groups, and we agree that the gravitational contributions to the decay rate can have important effects. However, as we explain in Sec. VII C, the problem of UV sensitivity is not just the coincidence between M_{Pl} and the flat point of the Higgs quartic: The Standard Model would be Planck sensitive even if M_{Pl} were 10^{100} GeV . Moreover, it is not correct to just use the effective potential to determine the bubble size; one really needs the full effective action, as we emphasize in Sec. VI.

A summary of some of the new perspectives provided in this paper is given in our conclusions, Sec. VIII.

II. TUNNELING IN QUANTUM MECHANICS

Much of our intuition for tunneling comes from one-dimensional quantum mechanics. Indeed, Gamow's 1928 calculation of the relation between half-life and the energy E of emitted α particles was seminal in establishing the validity of quantum mechanics [23]. So it is natural to start our discussion with this case. Gamow modeled the nuclear potential $V(x)$ as having a $\frac{1}{x}$ Coulomb tail and some kind of well for $x < a$, where the α particle is trapped. In 1D, the wave function of a state with energy E falls off exponentially between a and b by an amount given approximately by the WKB formula:

$$T(E) \equiv \frac{\psi_E(b)}{\psi_E(a)} \equiv e^{-W} \approx \exp \left[- \int_a^b dx \sqrt{2m(V(x) - E)} \right]. \quad (2.1)$$

Here, a and b are the turning points where $V(a) = V(b) = E$. It is of course quite logical that the decay rate should be proportional to how much of the wave function gets through the barrier, $\Gamma \sim |T(E)|^2$. However, if the particle is in an energy eigenstate, there is no time dependence, so it cannot decay. To go from $T(E)$ to Γ , a step often skipped, requires considerably more thought.

A simple picture often used to convert $T(E)$ to a decay rate depicts a particle with momentum $p = \sqrt{2mE}$ and velocity $v = \frac{p}{m}$ in the well hitting the barrier with a rate $\frac{v}{2a}$, each time tunneling through with probability given by the transmission coefficient, $|T(E)|^2$ (see e.g. [24]). With this logic, the decay rate is

$$\Gamma \sim \frac{p}{2am} \left| \frac{\psi_E(b)}{\psi_E(a)} \right|^2 \approx \frac{p}{2am} e^{-2W}. \quad (2.2)$$

Indeed, if one solves the Schrödinger equation numerically, one can see the wave function oscillate back and forth in the well; the largest flux leaks out during the times when the wave function is closest to the barrier. Figure 1 shows this exponential decay with time and the small oscillations. Snapshots of the wave function oscillating in the well are shown in Fig. 2.

A. Precise definition of the decay rate

To make the above formula more precise, we need an exact definition of the decay rate to which we can then look for approximations. A reasonable, physical definition of the decay rate of a system comes from $P_{\text{FV}}(t)$, the probability of finding a state ψ initially confined to a false-vacuum region (FV) in that same region after a time T :

$$P_{\text{FV}}(T) \equiv \int_{\text{FV}} dx |\psi(x, T)|^2. \quad (2.3)$$

We expect that for a decaying system the probability should fall exponentially:

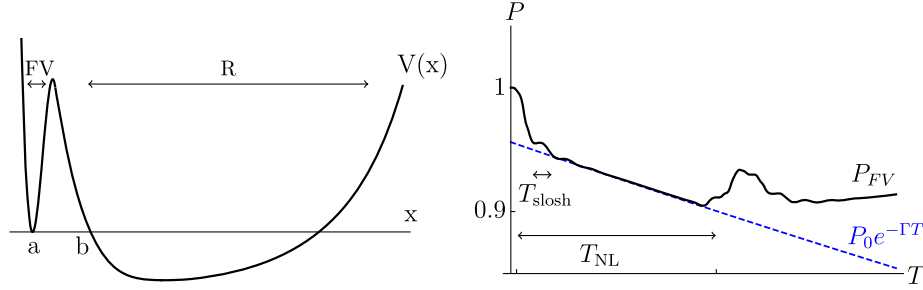


FIG. 1. Left panel: An example of a physical potential with a metastable region FV, a destination region R, and a barrier. We label the local minimum inside the FV region by a and the turning point by b [defined by $V(b) = V(a)$]. Right panel: The probability $P_{FV}(T)$ [see Eq. (2.3)] for this system (beginning in a Gaussian wavepacket centered at a) computed by numerically solving Schrödinger's equation. We see that the probability to find the particle in the false vacuum decays exponentially for intermediate times between the short time scale of sloshing inside the false vacuum and the long time scale on which the wave function begins to flow back into the false vacuum.

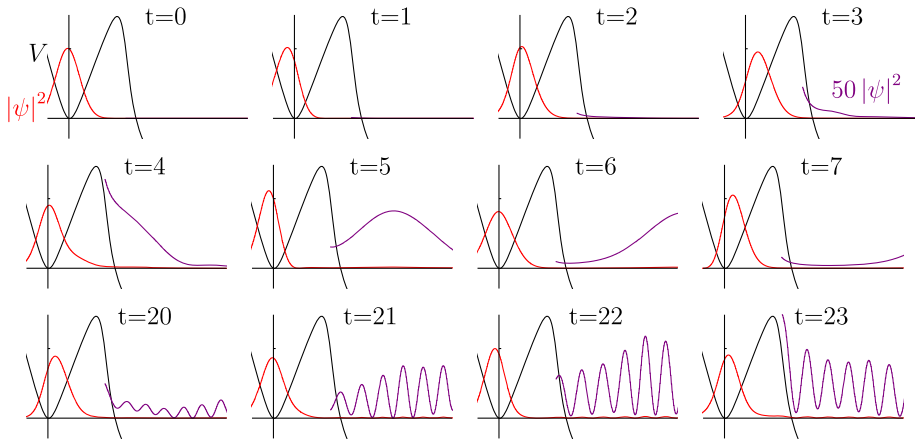


FIG. 2. The numerical evolution of a particle initially localized in the false vacuum. At each time step, the potential is shown (black), along with the probability $|\psi(x, t)|^2$ (red), and we also show the probability magnified by 50 (purple) so that we can see the small amount leaking through the barrier. By looking at the evolution of the wave function, we see the sloshing behavior near the false vacuum, associated with the initial Gaussian state not being an exact resonance. In the first two rows the central value of the wave function can be seen moving back and forth within the false-vacuum well. When it hits the right wall around times 3 and 4, the most wave-function amplitude escapes through the barrier. In the third row we have jumped ahead to see the nonlinear behavior when there is enough wave-function density in the outside region such that it is no longer simply flowing out.

$$P_{FV}(T) \sim e^{-\Gamma T}. \quad (2.4)$$

So we might define

$$\Gamma = -\frac{1}{P_{FV}} \frac{d}{dT} P_{FV}. \quad (2.5)$$

Equation (2.4) is of course not strictly true for all times T , and hence Γ defined in Eq. (2.5) is time dependent. For a 1D system, we can calculate $P_{FV}(T)$ numerically, as shown in Fig. 1. The plot of P_{FV} makes it clear what we mean physically by the “decay rate”; the probability falls exponentially for some particular time regime, and it is in this regime that Γ is meaningful.

To get a time-independent rate, we can average over the oscillations which occur with frequency $\sim \frac{p}{am}$. For this to

make sense, T cannot be too short: $T \gg T_{\text{slosh}} = \omega_a^{-1}$, where ω_a characterizes the frequency of oscillation within the false vacuum.¹ Moreover, T should also not be too long, for then the exponential decay will have significantly depleted the wave function and nonlinearities set in. One source of nonlinearities is from the decaying wave bouncing off the potential in the true-vacuum region and

¹For a parabolic well, the sloshing time is just the inverse of the classical oscillation frequency $\omega_a = \sqrt{\frac{1}{m} V''(a)}$. This is also the energy difference between excitations in the quantum system. For a square well, there is no classical oscillation, but T_{slosh} does not go to infinity. Because of the hard walls in the square well, there is still a finite T_{slosh} , related again to the energy difference of the two lowest modes.

returning to the false vacuum. Pooling these effects into a characteristic scale T_{NL} , we also require $T \ll T_{\text{NL}}$.

Thus, the physical decay rate is a phenomenon that happens on time scales $T_{\text{slosh}} \ll T \ll T_{\text{NL}}$; for these time scales we expect an exponential fall of $P_{\text{FV}}(T)$. Hence, Γ extracted from Eq. (2.5) is independent of T to the extent that $T_{\text{slosh}} \ll T \ll T_{\text{NL}}$ is satisfied. These two time limits are built into what is meant by a time-independent decay rate Γ ; they are not approximations we make to calculate Γ but rather limits under which Γ is even worth talking about. An “all-orders” formula for Γ *must* use these limits. Thus, a precise definition of the decay rate is

$$\Gamma \equiv - \lim_{\substack{T/T_{\text{NL}} \rightarrow 0 \\ T/T_{\text{slosh}} \rightarrow \infty}} \frac{1}{P_{\text{FV}}(T)} \frac{d}{dT} P_{\text{FV}}(T). \quad (2.6)$$

Finally, for systems that can decay in multiple different directions (e.g. a one-dimensional particle that can escape to the left or the right, or in multiple dimensions), we might want to know the decay rate to a particular region R (for instance, the region to the right of the barrier). Then we should define the partial decay width from the linear growth of the probability to find the particle in the region R, $P_{\text{R}}(T)$:

$$\Gamma_{\text{R}} = \lim_{\substack{T/T_{\text{NL}} \rightarrow 0 \\ T/T_{\text{slosh}} \rightarrow \infty}} \frac{1}{P_{\text{FV}}(T)} \frac{dP_{\text{R}}(T)}{dT} \quad (2.7)$$

Another way we could have intuitively derived the decay rate to any region R would be as the probability flux through the boundary of R. (If R is everything outside the false vacuum, then this would be the total flux into R, i.e. out through the boundary of the false-vacuum region.) The quantum mechanical flux is defined by

$$J_i(x, t) = \frac{1}{2im} (\psi^*(x, t) \partial_i \psi(x, t) - \psi(x, t) \partial_i \psi^*(x, t)). \quad (2.8)$$

Then we could define the decay rate as the fraction of probability flowing through the outward-pointing boundary ∂R , in the same time limits as above:

$$\Gamma_{\text{R}} \equiv - \lim_{\substack{T/T_{\text{NL}} \rightarrow 0 \\ T/T_{\text{slosh}} \rightarrow \infty}} \frac{1}{P_{\text{FV}}(T)} \int_{\partial\text{R}} dx_i J_i(x, T). \quad (2.9)$$

Because of the conservation equation ($\partial_t J_i = -\partial_i |\psi|^2$), this is exactly equivalent to Eq. (2.7).

Next, we need to be able to compute Γ in Eq. (2.7), either using the WKB approximation or with some other method.

B. Real-energy eigenstates and complex energy poles

The type of potentials under consideration, such as the one in Fig. 1, comprise a well region labeled FV, where the

particle is initially a barrier region B, between points a and b (to be specified precisely later), and an approximately free destination region R. For now, let us assume that the potential is constant in R and extends infinitely to the right, as $V(x)$ in Fig. 3.

A concrete example illustrating the points of this section is given in Appendix A. More details and alternative derivations can be found in [3,25–31].

Since the system extends infinitely to the right, there will be energy eigenstates $\phi_E(x)$ for any E . Most of these are approximately free (plane waves) confined to region R, with little support in the FV region. Some, however, do have large support in the FV region. These are the resonances. To be specific, we can define the resonant energies E as those whose probability in the FV region has a local maximum: $\partial_E P_{\text{FV}}[\phi_E] = 0$ [now the probability P_{FV} defined in Eq. (2.3) is viewed as depending on ϕ_E instead of on T , since energy eigenstates have time-independent probabilities]. In general, there will be a finite number of such resonance energies, $E_1 < E_2 < \dots < E_n < V_{\text{max}}$. Up to exponential corrections, these are the bound-state energies for a modified potential where FV is made absolutely stable by deforming the potential $V(x)$ [e.g., by setting $V(x) = V(c)$ for $x \geq c$, where $x = c$ is the location of the maximum height of the barrier].

Now, our initial state ψ cannot be an energy eigenstate, or else there would be no time dependence and $\Gamma = 0$. However, since the initial state is, by assumption, localized in the FV region, it can be written as a linear combination of bands of energy eigenstates close to each of the resonant energies E_i . Each of these bands will have a different characteristic decay rate Γ_i . We must assume $\Gamma_i \ll E_i$ so that the widths are narrow and the decay is exponential. As we will confirm, the higher energy states will decay much faster than the lower energy states since they have less barrier to penetrate: $\Gamma_1 \ll \Gamma_2 \ll \dots \ll \Gamma_n$.

As we said, we want to average over the sloshing times, $T \sim T_{\text{slosh}}^i = \omega_a^{-1}$. By the time $T \sim \Gamma_n^{-1} \gg T_{\text{slosh}}^n$, the highest energy components (around E_n) will have significantly leaked out of the well. Then, at each Γ_i threshold, another

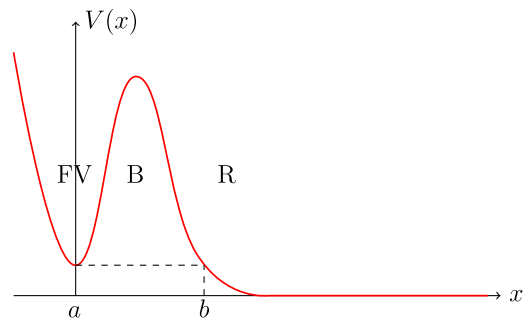


FIG. 3. Example of a potential that has a well region labeled FV and a barrier region B, and is constant in the region R which extends indefinitely to the right.

band of the original probability will have leaked out. Depending on the structure of the region R, these leaked components may even leak back in on the time scale $T \sim T_{\text{NL}}$. We can treat each band separately, so let us assume for simplicity that our original wave function only had support from the modes close to energy E_0 and write Γ for the width of this band.

The modes in the band near E_0 basically have the same form. They have the shape of the $E = E_0$ (approximate) bound state in the well region, decay exponentially through the barrier, and become free (plane waves) in region R.

In the FV region, they take the form

$$\phi_E(x) = \frac{1}{N(E)} f(E, x) \quad (2.10)$$

where $f(E, x)$ is a well-behaved function in E and x , and $N(E)$ has zeros exponentially close to the resonant energies E_n . We are focusing on the lowest resonant energy, which has corresponding zeros $E_0 \pm \frac{i}{2}\Gamma_0$.

To connect Γ in Eq. (2.6) to this pole, let us decompose our decaying wave function $\psi(x, t)$, initially localized in the well, into energy eigenstates. We define

$$\rho(E) = \int_0^\infty dy \phi_E^*(y) \psi(y), \quad (2.11)$$

where $\psi(y) \equiv \psi(y, t=0)$. The time dependence of our wave function $\psi(x, t)$ is given by

$$\begin{aligned} \psi(x, t) &= \int dE \rho(E) \phi_E(x) e^{-iEt} \\ &= \int dy \psi(y) \int dE \frac{f(E, x) f^*(E, y)}{|N(E)|^2} e^{-iEt}. \end{aligned} \quad (2.12)$$

Closing the E contour in the fourth quadrant (for convergence), and looking at times $E_0^{-1} \ll t \ll \Gamma_0^{-1}$, we have

$$\begin{aligned} \psi(x, t) &\approx \int dy \psi(y) 2\pi i \text{Res} \left[\frac{f(E, x) f^*(E, y)}{|N(E)|^2} e^{-iEt}; \right. \\ &\quad \left. E = E_0 - \frac{i}{2}\Gamma_0 \right] \\ &= g\left(E_0 - \frac{i}{2}\Gamma_0, x\right) e^{-iE_0 t} e^{-\frac{\Gamma_0}{2}t} \end{aligned} \quad (2.13)$$

for some well-behaved function $g(E, x)$. Corrections to this formula are all suppressed by $\Gamma_0 t \ll 1$. Applying the definition of the decay rate, Eq. (2.6) gives

$$\begin{aligned} -\frac{d}{dT} \log P_{\text{FV}}(T) &= -\frac{d}{dT} \log \int_{\text{FV}} dx |\psi|^2 \\ &= -\frac{d}{dT} \log [\text{const} \times e^{-\Gamma_0 T}] = \Gamma_0. \end{aligned} \quad (2.14)$$

Thus Γ_0 , defined by the first pole of $N(E)$, $E = E_0 - \frac{i}{2}\Gamma_0$, is indeed the decay rate Γ . That is, the decay rate, defined

physically in Eq. (2.6), is given by twice the imaginary part of the complex pole. For potentials of this form, this leads us to a shortcut; we can simply calculate the poles of $\frac{1}{N(E)}$ directly, without ever explicitly time-evolving any states.

To be clear, real-energy eigenstates never blow up. However, nothing stops us from finding solutions to the Schrödinger equation with complex energies. In doing so, we find that for certain complex energies, the wave function does blow up. These poles should be close to real energy E_0 , with a small excursion of size $\Gamma \ll E_0$ into the complex plane. Actually, there will be two poles for each E_0 , one for positive and one for negative Γ . Since the Hamiltonian is Hermitian on bound states, these states cannot have the same boundary conditions as for real-energy eigenstates. Normally, energy eigenstates with real energies have a probability around each point which does not change with time, so there can be no outgoing or incoming flux. Thus, nonzero flux corresponds to complex energies. The $\Gamma < 0$ case corresponds to incoming boundary conditions: Flux goes from region R into the FV region, and correspondingly, $P_{\text{FV}}(T)$ will grow with time. For $\Gamma > 0$, the flux goes from FV into R. These are outgoing radiating (Gamow-Siegert) boundary conditions, the situation we are interested in. The example in Appendix A shows more directly the connection between radiating boundary conditions and the complex zeros of the normalization.

Note that the outgoing-only wave approximation is equivalent to removing the backreaction, or equivalently taking $T/T_{\text{NL}} \rightarrow 0$. This is the same limit required in Eq. (2.7) to make the decay rate well defined (time independent). It is reassuring that the relevant time scale plays a role in the analysis.

With outgoing boundary conditions, the energy is $E = E_0 - \frac{i}{2}\Gamma$ with $\Gamma > 0$, and the momentum in the region where $V = 0$ is $p = p_0 - \frac{i}{2}\gamma$, with $\gamma = \frac{m\Gamma}{p_0}$ and $p_0 = \sqrt{2mE_0}[1 + \mathcal{O}(\Gamma/E_0)]$. Writing $\phi_E(x, t) = \frac{1}{N} e^{-iEt + ipx}$ in the region R (where we are assuming $V = 0$), the rate can be computed by flux conservation,

$$\partial_t(\psi^* \psi) = \frac{i}{2m} \partial_x(\psi^* \partial_x \psi - \psi \partial_x \psi^*), \quad (2.15)$$

which holds for any solution ψ to the Schrödinger equation. Integrating this from 0 to an arbitrary point b for the energy eigenstate ϕ_E with outgoing boundary conditions, we then get

$$\Gamma = \frac{p_b}{m} \frac{|\phi_E(b)|^2}{\int_0^b dx |\phi_E(x)|^2} \quad (2.16)$$

where $p_b = -i \frac{\partial_x \phi_E(b)}{\phi_E(b)}$ for a plane wave, and more generally $p_b = -\frac{i}{2} \frac{\phi_E^* \partial_x \phi_E - \phi_E \partial_x \phi_E^*}{\phi_E^* \phi_E} \Big|_{x=b}$.

The expression for the decay rate in Eq. (2.16) is accurate up to exponentially small corrections. In it, b can be any point *at all*. Indeed, Eq. (2.16) is independent of the choice of b , as can be seen by taking the derivative and applying Schrödinger's equation, and it is independent of the choice of t because it is written in terms of spatial wave functions alone.

For Eq. (2.16) to be useful, we would like to be able to use it for solutions with real energies for normalizable resonance modes rather than complex energy eigenstates with outgoing boundary conditions. After all, if we already know the complex energy, then we know the rate. Since the complex energy solution is exponentially close to the real-energy solution in the FV region and for most of the barrier region (except near $x \approx b$), the integral $\int_0^b |\phi_E(x)|^2$ in the denominator should be about the same if ϕ_E is the real or complex energy solution. The numerator, on the other hand, involves the value of the wave function at b . Its value for the real and complex-energy solutions may differ by a factor of order one. This is because the complex energy solution has an exponentially growing component in the barrier, which can become of the same order as the exponentially decaying one at $x = b$. Thus, while one can certainly use Eq. (2.16) with a real eigenfunction to approximate the exact answer, some precision will unfortunately be lost in doing so. We discuss using the WKB approximation to the energy eigenstates next and defer an explicit example to Appendix A.

C. WKB approximation

Once we have a formula like Eq. (2.16) which depends on the values of a wave function, we need to solve Schrödinger's equation. If an analytic solution is not available, we may want to approximate Eq. (2.16) with the WKB expansion. The WKB approximation tells us that

$$\phi_E(x) = A \frac{1}{\sqrt{|p(x)|}} \exp \left[\frac{i}{\hbar} \int_a^x p(y) dy \right] (1 + \mathcal{O}(\hbar)), \quad (2.17)$$

where $p(x) = \sqrt{2m(E_0 - V(x))}$ and A is an x -independent normalization constant that will drop out of Eq. (2.16). The lower limit of integration is chosen to be a for convenience; changing it to something else will only change the normalization, which can then be absorbed into A .

At leading order we ignore the p prefactor and keep only the exponential. Then Eq. (2.16) gives

$$\Gamma^{\text{LO}} = \text{const} \times e^{-2 \int |p(x)| dx}, \quad (2.18)$$

where the integral is taken over the region where $V(x) > E_0$ [so that $p(x)$ is imaginary]. At NLO we keep the prefactor also, giving

$$\Gamma^{\text{NLO}} = \frac{1}{m} \frac{\exp \left[-\frac{2}{\hbar} \int_a^b |p(x)| dx \right]}{\int_0^a \frac{dx}{p(x)} + \int_a^b \frac{dx}{p(x)} \exp \left[-\frac{2}{\hbar} \int_a^x |p(y)| dy \right]} \quad (2.19)$$

where a and b are the classical turning points.

The first factor in the denominator is exactly the classical period of oscillation around the false vacuum: $T_{\text{FV}} = \int \frac{m}{p} dx$. For a roughly constant potential in the region from 0 to a , this is just $T_{\text{FV}} = \frac{a}{v}$ where $v = \frac{p}{m}$ is the velocity. The second factor in the denominator is exponentially smaller in the limit $\hbar \rightarrow 0$, so we can drop it even at NLO. Thus,

$$\Gamma^{\text{NLO}} = \frac{p}{am} \exp \left[-\frac{2}{\hbar} \int_a^b |p(x)| dx \right]. \quad (2.20)$$

This is close to Eq. (2.2), but differs by a factor of 2.

Of course we had no right to expect the two formulas to agree exactly: First, Eq. (2.2) is based on an imprecise semiclassical argument, and second, Eq. (2.20) uses the WKB approximation, neglecting a careful treatment of turning points, and approximates Eq. (2.16) with real-energy eigenstates, which is also not a controlled approximation. In [4], the WKB approximation is used in a formula like Eq. (2.20) for a cubic potential. They find that at NLO the prefactor differs from a presumably more accurate result using the potential-deformation method (see Sec. III below), by a factor of $\frac{\epsilon}{2} \approx 1.4$.

This is not to say that WKB cannot be used to compute tunneling rates precisely. It can. For example, in [32] the WKB method was used to compute the rate from the complex resonant energies for the quartic potential to N^4LO . The same rate was computed to NLO using the potential-deformation method in [33]. The two results agree exactly to the order at which they can be compared (NLO).

D. WKB in multiple dimensions

One might wonder what happens to the above Schrödinger equations, especially Eqs. (2.16) and (2.20), in multiple dimensions. The all-orders formula Eq. (2.16) generalizes naturally enough to

$$\Gamma = \frac{1}{m} \frac{\int_{\Sigma} db \cdot p_b |\phi_E(b)|^2}{\int_{\text{FV}} dx |\phi_E(x)|^2}. \quad (2.21)$$

The WKB approximation in multiple dimensions is more complicated, unfortunately. At an intuitive level, one would very much like to simply integrate over all paths through configuration space, and for each path apply the 1D WKB. In other words, this simply says that the system can decay along any path through the barrier; for each path we apply the 1D WKB formula and then we integrate over all the paths.

This intuitive picture is unfortunately difficult to prove precisely (an extended discussion is provided by Banks,

Bender, and Wu [9]). The problem is that WKB is attempting to approximate the wave function, which takes only a single value at each position; there is no sense in summing over all possible values it would take if we follow all possible paths to each point. Altogether, when we want to be precise, it is easier to apply the semiclassical approximation (which is the approximation WKB performs) in the path integral rather than trying to use multidimensional WKB.

However, the intuition from WKB is not useless. Because the trouble has to do with the sum over paths, one might expect the leading exponential behavior predicted by WKB—that which is determined solely by the dominant path through the barrier and knows nothing of the other paths—to be correct. Keeping only the leading exponentials, this says

$$\Gamma \sim \int \tilde{\mathcal{D}}x e^{-2 \int ds \sqrt{2V(x(s))}}, \quad (2.22)$$

where the path integral here integrates over paths but not over parametrizations of those paths; in other words, we only include in $\tilde{\mathcal{D}}x$ paths $x(s)$ with a path-length normalization $|\frac{dx_i}{ds}|^2 = 1$.

For any path $x(s)$, the WKB exponential is exactly the same as the minimal classical Euclidean action over all parametrizations $s(t)$:

$$\int ds \sqrt{2V(x(s))} = \min_{s(t)} \int dt \left[\frac{1}{2} \dot{x}^2 - V(x) \right]. \quad (2.23)$$

This can be seen because the minimum action path conserves energy; $E = \frac{1}{2} \dot{x}^2 + V(x)$ is constant. Assuming $V = 0$ at the end point, this means $\dot{x} = \sqrt{2V}$, and the minimum action is equal to $\int dt 2V(x)$. Changing variables from dt to ds gives $\int ds \sqrt{2V}$. Note that both sides of Eq. (2.23) only integrate over the path from the false vacuum to the barrier; to include the return journey, one adds a factor of 2.

This means that (again keeping only the dominant exponentials) integrating over the parametrizations of the Euclidean action along a single path gives the WKB factor along that path:

$$\int \mathcal{D}s(t) e^{-S_E[x(s(t))]} \sim e^{-2 \int ds \sqrt{2V(x(s))}}, \quad (2.24)$$

where the left side is integrated over all paths from false vacuum back to false vacuum which cross the barrier, and the right side integrates from the false vacuum to the turning point. Now we can remove the awkward restriction on the path measure in Eq. (2.22):

$$\Gamma \sim \int \mathcal{D}x e^{-S_E[x]}, \quad (2.25)$$

which is indeed the correct equation at leading exponential order according to the more precise path-integral derivations (see Secs. III and IV below), as long as one allows the \sim to suppress some sort of restriction to the bounce saddle point only.

Thus, the picture of using WKB and simply integrating over paths through the barrier (or simply taking the least-resistance path) does indeed give the correct leading-order decay rate. This “through the barrier” restriction causes the integral in Eq. (2.25) to be dominated by the bounce and not the constant false-vacuum solution (see Sec. III).

This intuition is useful since it says that for a given multidimensional potential, one can get a physical intuition for the size of the barrier by studying the potential along the least-resistance path $V_{1D}(s) \equiv V(x(s))$, where $x(s)$ is the least-resistance path which is the same as the Euclidean bounce. The leading exponential decay rate for the multidimensional problem will then be the same as it would be for the 1D problem $V(s)$.

As a side note, it is important that the integration variable s in WKB be a path-length parametrization. In field theory (cf. Sec. V below), we usually parametrize the path through field space with the Euclidean time τ . But if we use τ as is, then the WKB factor $\int d\tau \sqrt{2V}$ has the wrong measure. So we must first convert to a path-length parametrization:

$$\frac{ds}{d\tau} = \sqrt{\sum_i \left(\frac{dx_i}{d\tau} \right)^2}. \quad (2.26)$$

This is discussed further in Sec. V B.

E. Summary

The first goal was to give a precise definition of the decay rate and to isolate the conditions under which Γ is well defined. We did this in Sec. II A. The next goal was to show how the transmission coefficient, which is what the WKB can be used to compute, is related to Γ . To do that, we needed to discuss the time evolution of a wave function ψ . We found, from decomposing ψ into energy eigenstates, that the rate is encoded in the zeros of the normalization of modes near resonant energies E_0 . This normalization has a pole at complex energies. The pole whose imaginary part gives the rate is associated with outgoing-radiation boundary conditions. Using flux conservation, this imaginary part can be related to the energy eigenstate wave function which is then approximated with WKB. Although all these steps are presumably well known, and included in various forms in various treatments, we nevertheless thought it could be helpful to have this whole story in one place.

Appendix A explains a concrete example, where the (real) energy eigenstates are solved explicitly, and the connection between complex energies, poles in the normalization of the wave functions, and outgoing boundary conditions can be seen explicitly.

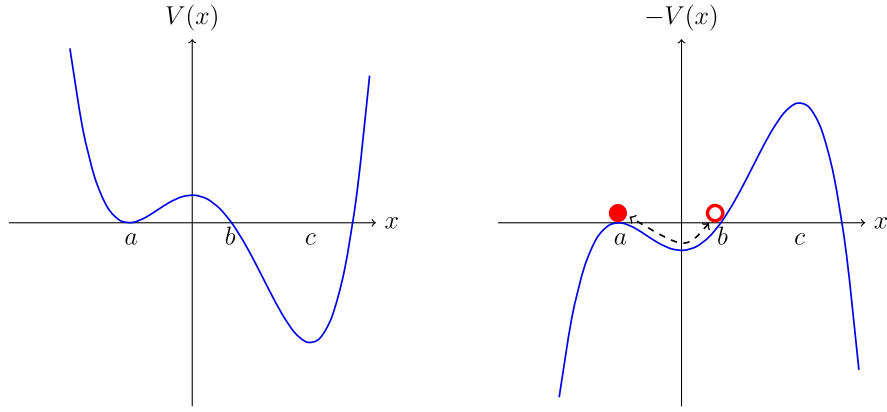


FIG. 4. Left panel: Generic potential with a false and true vacuum. Right panel: The inverted potential. The stationary path $\bar{x}(\tau)$ is the solution to the equations of motion of a ball rolling down the inverted potential with boundary conditions $x(0) = x_i$ and $x(T) = x_f$.

III. POTENTIAL-DEFORMATION METHOD

The methods of the previous section rely on solving Schrödinger's equation, which is not practical in a many-dimensional case (such as field theory). An alternative approach to calculating tunneling rates, which generalizes more easily to higher dimensions, works with the path integral directly [1]. We first review this approach and point out some of its more curious aspects. Then in Secs. III B–III B 3 we provide more details of particularly subtle points. More details of the mathematics of this method can be found in [5,6,34–36].

A. Overview

The starting point of the calculation in [1] is the relation

$$Z \equiv \langle x_f | e^{-HT} | x_i \rangle = \int_{x(0)=x_i}^{x(T)=x_f} \mathcal{D}x e^{-S_E[x]}, \quad (3.1)$$

where the right-hand side is the path integral using the Euclidean action $S_E[x]$. By inserting a complete set of energy eigenstates, the matrix element can be written as

$$Z = \sum_E e^{-ET} \phi_E(x_i) \phi_E^*(x_f). \quad (3.2)$$

Then we see that the lowest energy can be deduced from

$$E_0 = -\lim_{T \rightarrow \infty} \frac{1}{T} \ln Z. \quad (3.3)$$

Roughly speaking, we expect that when there is a decay, E_0 will have an imaginary part corresponding to the decay rate, so²

²Note that there will be a sign ambiguity in the evaluation of Eq. (3.4), as we will see later in this section. The calculation should always be done so that $\Gamma > 0$, which corresponds to the physical decay rate.

$$\frac{\Gamma}{2} = \text{Im} \lim_{T \rightarrow \infty} \frac{1}{T} \ln Z. \quad (3.4)$$

There are many ways to connect the imaginary part of an energy to a decay rate, but the connection is not automatic. For example, in Sec. II B we found the decay rate for a metastable system to be the imaginary part of an eigenstate of the Hamiltonian with (unphysical) Gamow-Siegert radiative boundary conditions. For normalizable modes of a Hermitian Hamiltonian, all the energies including E_0 are real. For physical potentials, which are bounded from below, the energies and Euclidean action are bounded from below as well. Correspondingly, Z is manifestly real. Hence, Eq. (3.4) must be defined in a much more careful manner.

Consider the asymmetric double-well potential in Fig. 4. One might hope that by taking particular boundary conditions (choice of x_i and x_f) we can extract a metastable or resonance energy whose imaginary part gives the decay rate. However, the points x_i and x_f only contribute through the wave-function factors $\phi_E(x_i)$ and $\phi_E^*(x_f)$ in Eq. (3.2), which do not contribute to E_0 . In order to get an imaginary part, then, we must do something more tortuous.

Since the path integral is complicated, let us simplify things by first approximating it using the saddle-point approximation. The path integral can be approximated by summing over stationary points of the Euclidean action. For each stationary point, that is, for each solution $\bar{x}(\tau)$ to the Euclidean equations of motion, the saddle-point approximation of the path integral around \bar{x} evaluates to

$$\mathcal{I}_{\bar{x}} \equiv \frac{N}{\sqrt{\det[-\partial_t^2 + V''(\bar{x})]}} e^{-\frac{1}{\hbar} S_E(\bar{x})} (1 + \mathcal{O}(\hbar)) \quad (3.5)$$

where we have put the \hbar back in for clarity, and N is some constant related to the normalization of the path integral. The stationary paths $\bar{x}(\tau)$ are solutions to the equations of motions for a ball rolling down a hill described by the

inverted potential $-V(x)$ with the boundary conditions from the path integral: $x(0) = x_i$ and $x(T) = x_f$ (see Fig. 4).

There are a range of stationary paths for this system dependent on what we choose the boundary conditions to be. For $x_i = x_f = c$ there is a solution to the Euclidean equations of motion with $\bar{x}(\tau) = c$. This is the path labeled “TV static” in Fig. 5. This has an action given by $S_E[\bar{x}] = V(c)T$. The integral over Gaussian fluctuations around this solution produces $\mathcal{I} \sim \exp(-E_c T)$, as explained in [1], where $E_c = V(c) + \frac{1}{2}\sqrt{V''(c)}$ is the ground-state energy of a harmonic oscillator using the quadratic approximation to the potential near $x = c$. Thus, the $T \rightarrow \infty$ limit produces the correct approximate ground-state energy $E_0 = E_c$ for $x_i = x_f = c$, as expected.

Now say we take x_i and x_f to be arbitrary (not at c). The Euclidean equations of motion with boundary conditions $x(0) = x_i$ and $x(T) = x_f$ can always be solved by a solution starting at x_i with exactly enough initial velocity to get to the top of the hill and stay there for nearly time T , and then roll to x_f . This path is shown as the path labeled “generic shot” in Fig. 5. This path has nearly the same Euclidean action as the TV static path, and matches it exactly as $T \rightarrow \infty$. Thus, we can indeed choose any points x_i and x_f , and the true ground-state energy E_0 results from the $T \rightarrow \infty$ limit.

Now consider $x_i = x_f = a$. There is still a solution to the Euclidean equations of motion which stays at $x = c$ for most of the time (the path labeled the “shot” in Fig. 5) with $\mathcal{I} \sim \exp(-E_c T)$. With $x_i = x_f = a$ there are actually more solutions. The one labeled “FV static” stays at $x = a$ for all times. It has Euclidean action $V(a)T$ and saddle-point approximation $\mathcal{I} \sim \exp(-E_a T)$, where just like for E_c , E_a has corrections due to oscillations around the $x = a$ minima. In addition, when $x_i = x_f = a$ there is also an exact instanton solution. This solution starts very slowly from a . Since the potential is flat at a , it stays near a for a long time, then rather quickly rolls up to b and back, and then stays near a again for a long time. This is the “bounce” in Fig. 5. Because it spends most of the time near a , its

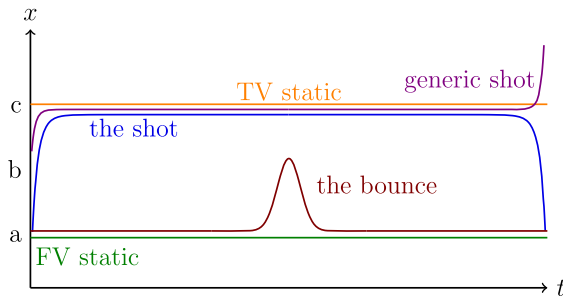


FIG. 5. Different solutions to the Euclidean equations of motion for the asymmetric double well.

saddle point differs from the FV saddle by a finite (T -independent) amount: $\mathcal{I} = K \exp(-E_a T - S_{\text{bounce}})$, for some imaginary coefficient K . The finite action is $S_{\text{bounce}} = \int_a^b dx \sqrt{2mV(x)}$, where the integral is between a and the turning point b on the other side of the barrier. This action is positive and therefore larger than the action for the shot, which also starts and ends at a in time T . The bounce only moves for a finite time, which is why it is called an instanton.

So when $x_i = x_f = a$, there are contributions to the path integral from the shot, the FV static, and the unique bounce which is an exact solution to the Euclidean equations of motion. There are also approximate solutions where the bounce is translated or multiple bounces are sewn together. These have actions which are exponentially close to the bounce action and therefore contribute a large amount to the path integral even if they are not exact stationary points. Summing all the saddle points and approximate saddle points, the result is

$$Z = \langle A | e^{-HT} | A \rangle \sim \exp(-E_c T) + \exp(-E_a T) + \exp(-E_a T + K e^{-S_{\text{bounce}} T}) + \dots \quad (3.6)$$

where K is a NLO constant arising from Gaussian integrations around the bounce [1].

One might then argue (see Coleman’s discussion in [11]) that the bounces are the only thing with an imaginary part, so we can keep them when computing the imaginary part in Eq. (3.4), giving us

$$\frac{\Gamma}{2} = \text{Im} K e^{-S_{\text{bounce}}}. \quad (3.7)$$

But this argument is very precarious; we know for a fact that when computed exactly, Z is real and the imaginary part is exactly 0. The imaginary bounce contribution is exactly canceled by subdominant corrections to the true-vacuum saddle point. This cancellation can be seen in the toy examples discussed in Sec. III B.

In the following sections we will discuss some subtle points about saddle-point approximations, analytic continuation, and deformation of the contour of integration that will lead us to an expression like Eq. (3.7). Briefly, the decay rate is actually calculated by modifying the path integral to be along a different contour of integration, the contour of steepest descent through the FV. Integrating along this contour misses the shot solution, allowing the FV path to dominate. The imaginary part along this contour is the same as $\frac{1}{2}$ of the imaginary part along the steepest-descent contour passing through the bounce saddle point. This tells us that the decay rate associated with the false vacuum of the potential in Fig. 4 is given by

$$\frac{\Gamma}{2} = \frac{1}{2} \text{Im} K e^{-S_{\text{bounce}}} \quad (3.8)$$

which is $\frac{1}{2}$ of the naive result given in Eq. (3.7).

We now turn to a careful explanation of how the contour deformation and saddle-point approximation is done, and to what extent the result is the same as the one given by the analytic continuation of the path integral associated with deforming the potential.

B. Analytic continuation, steepest-descent contours, saddle points, and imaginary parts

The main goal of this section is to explain a mathematically consistent procedure for getting an imaginary part, presumably connected to the decay rate of a metastable state, out of a real path integral. This section is based, to a large extent, on [36] with insights from [5,6,34,35]. In contrast to these references, we also consider tunneling in physical bounded potentials, which leads to a more nuanced picture of the origin of the imaginary part.

The final result, which we are trying to justify, is that **the imaginary part we want comes from applying the method of steepest descent to the Euclidean action along a family of paths passing through the FV saddle point**. Much of the relevant mathematics can be understood most clearly by reducing the calculation to a one-dimensional integral along only this family of paths. Parametrizing the family by a parameter z , with $z = 0$ corresponding to the false vacuum, $z = 1$ the bounce, and $z > 1$ going towards the shot, we can compute directly $S(z)$ for a given potential (see Fig. 6). The part of the path integral of interest is then a 1D integral,

$$Z_C = \int_C dz e^{-\frac{1}{\hbar} S(z)}, \quad (3.9)$$

with C some integration contour, in this case simply the real line. Eventually, we will deform C to some other contour in order to calculate the decay rate.

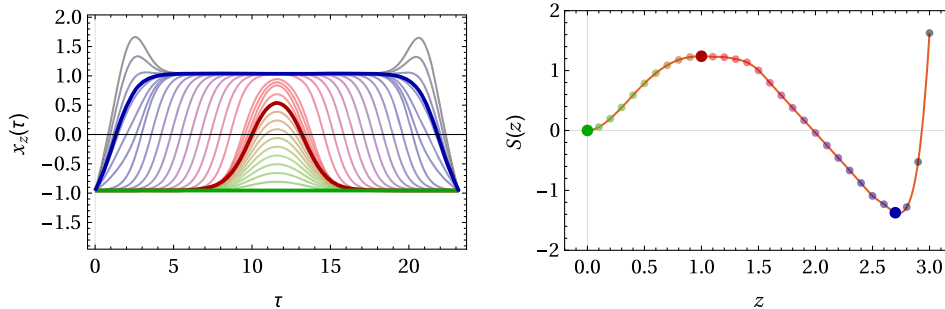


FIG. 6. It is helpful to study the path integral along a one-parameter family of paths passing from the static FV through the bounce and to the shot. These paths $x_z(\tau)$ are illustrated on the left (found numerically using the method in Appendix C) for the potential $V(x) = -\frac{x^2}{12} - \frac{x^2}{2} + \frac{x^4}{4}$, and their actions $S(z)$ are computed numerically and shown on the right. For the examples we consider, along families of paths like this, the action $S[z] = \int d\tau [\frac{1}{2}(\partial_\tau x_z)^2 + V(x_z)]$ looks qualitatively like $V(z)$.

First, in Sec. III B 1, we discuss the saddle-point approximation, in general, and explain why only a subset of the stationary points of the action contribute to the decay rate. Then we explain in a concrete example how the integral as a whole can be real even though a subdominant saddle point is imaginary [this is what happened in the path integral in Eq. (3.6)].

We unravel the origin of the imaginary parts by considering different types of potentials. One can have potentials in which the destination region of tunneling is unbounded from below, like $V(x) = \frac{x^2}{2} - \frac{x^4}{4}$. These examples are discussed in most textbook treatments [5–7]. For such cases, the energy spectrum is unbounded from below, and the path integral is formally infinite. One can produce the tunneling rate by deforming the potential through a parameter g to the potential where $g = 1$ is the original case of interest and $g < 0$ makes the path-integral convergent. As we will see, analytically continuing back to $g = 1$ corresponds to changing the integration contour into the complex z plane, giving Z a well-defined imaginary part.

Potentials in actual physical systems are necessarily bounded from below. One would naively expect that the same logic from the unbounded potentials should apply. However, as we will see, analytic continuation of the potential cannot produce an imaginary part in the path integral because Z is convergent along the real axis for any g . In particular, for physical potentials, applying Eq. (3.3) necessarily gives the *real* ground-state energy of the system.

The correct procedure, which applies for all types of potentials exhibiting tunneling, is to compute the path integral along the steepest-descent contour through the FV saddle. Along this contour, Z is complex, and its imaginary part is equal to $\frac{1}{2}$ the sum over the bounces, just as in the standard formula. However, this understanding gives little explanation for why this procedure should always give the decay rate or how to calculate the rate outside of the saddle-point approximation. Those questions are answered by the alternative method presented in Sec. IV.

1. Dominant and subdominant saddle points

The saddle-point approximation is used for arbitrary complex exponential integrals, like that in Eq. (3.9). To apply the approximation, we first ignore the actual contour C and instead focus on the complex saddle points of S , which we label s_1, s_2, \dots, s_n . Through each of these points, we draw the steepest-descent contour C_i , which is defined intuitively by simply moving away from s_i in the direction which increases the real part of S as quickly as possible. These contours are called “steepest descent” because the magnitude of the integrand is rapidly diminishing along the contour away from the saddle point, and thus the value of the integral can be approximated by its behavior near the maximum at s_i .

Along a given steepest-descent contour, we can approximate the integral by expanding S around s_i :

$$\mathcal{J}_i \equiv \int_{C_i} dz e^{-\frac{1}{\hbar} S(z)} \sim \int_{C_i} dz e^{-\frac{1}{\hbar} S(s_i) - \frac{1}{2\hbar} S''(s_i)(z-s_i)^2 + \dots} \quad (3.10)$$

$$\sim \sqrt{\frac{2\pi\hbar}{S''(s_i)}} e^{-\frac{1}{\hbar} S(s_i)} (1 + \mathcal{O}(\hbar)) \equiv \mathcal{I}_i. \quad (3.11)$$

Here, \mathcal{I}_i is an approximation to \mathcal{J}_i , approaching it exactly in the $\hbar \rightarrow 0$ limit. The subleading corrections in “ $\mathcal{O}(\hbar)$ ” can be calculated in an asymptotic series in \hbar . Because the series is asymptotic, summing the series will not reproduce \mathcal{J}_i exactly. It will only produce \mathcal{J}_i up to terms exponentially suppressed in $\frac{1}{\hbar}$.

We can then make a plot of the complex z plane, marking the saddle points of S as well as their steepest-descent contours, as shown in Fig. 7. Suppose the integral we want to compute is along the original contour C . We can deform the contour C into a sum of steepest-descent contours C_i ; in the example of Fig. 7, this would be

$$C = C_1 + C_2. \quad (3.12)$$

There are no poles, so the deformation is allowed as long as the end points remain the convergent regions indicated by the arcs in Fig. 7. Note that although there are three saddle points in this example, only two contribute because of the contour.

The integral along C is the sum of the integrals along these steepest-descent contours, yielding

$$Z = \mathcal{J}_1 + \mathcal{J}_2. \quad (3.13)$$

Note that the fact that some or all of the s_i are complex does not matter; one still includes those s_i whose contours are involved in the sum, regardless of whether the saddle point itself is real or not [37].

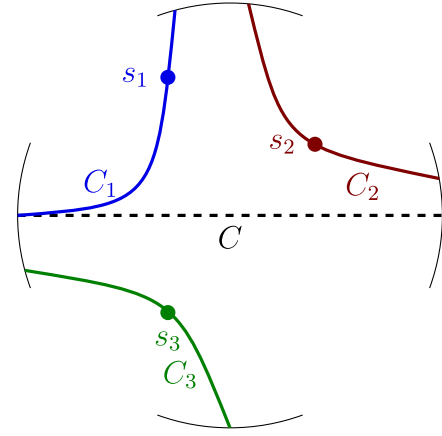


FIG. 7. An example of the complex z plane with the critical points of S marked as dots. Each saddle point s_i has a steepest-descent contour C_i passing through it. Black arcs around the edges note the directions in the x plane in which the exponent $-S$ goes to $-\infty$; the integral only converges along contours which start and end in these directions.

Now we can use the saddle-point approximation along each contour. We write the saddle-point approximation to \mathcal{J}_i as \mathcal{I}_i , so that

$$Z \sim \mathcal{I}_1 + \mathcal{I}_2, \quad (3.14)$$

where \sim indicates that corrections are exponentially small.

However, note that each \mathcal{I}_i approximates an exact contour integral \mathcal{J}_i up to exponentially suppressed terms, and the \mathcal{I}_i can be exponentially different. In other words, one of the terms in Eq. (3.13) (say \mathcal{J}_1) is exponentially larger than the rest. So if we are going to perform an expansion which is accurate up to exponentially small corrections, we cannot keep the subdominant terms. At the level of Eq. (3.13), we write this as

$$Z = \boxed{\mathcal{J}_1} + \mathcal{J}_2 \quad (3.15)$$

to indicate that while the equation is exact at this level, the boxed term is exponentially dominant so, when approximating, the second term is meaningless. Unfortunately, it may be \mathcal{J}_2 that has the imaginary part.

The problem of subdominant imaginary contributions is perhaps easiest to appreciate through an example. Suppose we have the following function S :

$$S(z) = -\frac{z^2}{2} + \frac{z^4}{4}, \quad (3.16)$$

and we want to integrate as in Eq. (3.9) along the real line. This function has saddle points at $z = -1, 0, 1$, with approximations

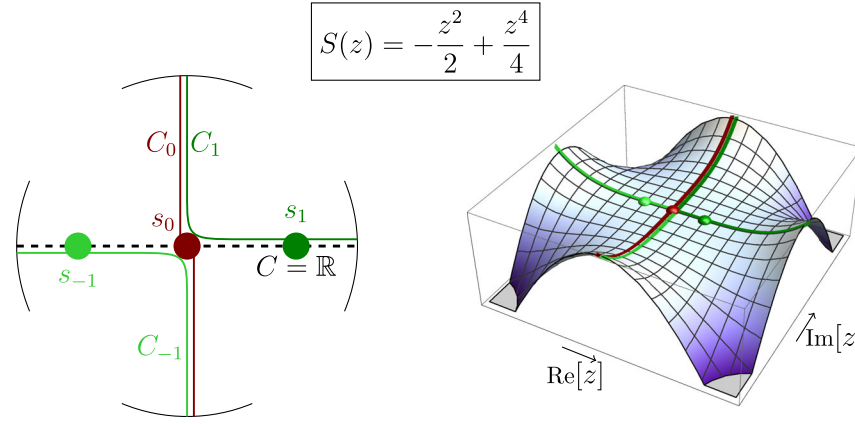


FIG. 8. The saddle points $\{-1, 0, 1\}$ and associated steepest-descent contours for $S(z) = -\frac{z^2}{2} + \frac{z^4}{4}$. Left panel: The real line (dotted) can be written as a sum of all three contours $\mathbb{R} = C_{-1} + C_0 + C_1$. Actually, there is an ambiguity for this action (it is on a Stokes line); the complex conjugate contours are equally valid. Right panel: A plot of $\text{Re}[S(z)]$ in the complex plane where we can clearly see the lines of steepest descent through each saddle point.

$$\begin{aligned} \mathcal{I}_{-1} &= \sqrt{\pi\hbar} \exp\left(\frac{1}{4\hbar}\right) & \mathcal{I}_0 &= \sqrt{-2\pi\hbar} \\ \mathcal{I}_1 &= \sqrt{\pi\hbar} \exp\left(\frac{1}{4\hbar}\right). \end{aligned} \quad (3.17)$$

Around $z = 0$, the quadratic action $S(z) = -\frac{z^2}{2}$ has an increasing real part along the imaginary axis. Around $z = 1$, $S(z) = -\frac{1}{4} + (z-1)^2$, which has an increasing real part along the real axis. Thus, going from $z = 1$, the

steepest-descent contour moves along the real axis, until it hits $z = 0$, where it must turn to either the positive or negative imaginary direction. This ambiguity (this action is said to be on a Stokes line) is easily resolved by giving the action a small imaginary part. For one choice, the steepest-descent contours are sketched in Fig. 8. The steepest-descent contours are sometimes called Lefschetz thimbles. For the other choice, the thimbles would be the complex conjugates of those in Fig. 8 (see Fig. 9 below for another example which shows the conjugate contours).

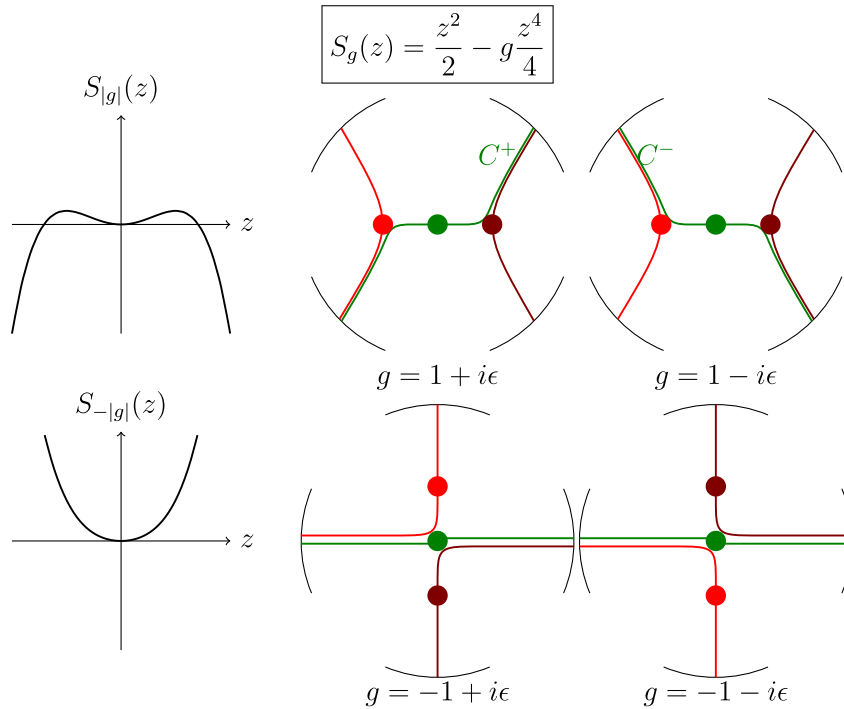


FIG. 9. The saddle points and steepest-descent contours for $S_g(z) = \frac{z^2}{2} - g\frac{z^4}{4}$ in Eq. (3.20). For $g = 1$, integrating along the real axis is divergent (as indicated by the lack of arcs at $z = \pm\infty$). For $g = -1$, the FV contour (green line) falls along the real axis. Rotating g back from -1 to 1 , the FV contour remains convergent but depends on whether one rotates g clockwise or counterclockwise in the complex plane.

We see from Fig. 8 that in this case, the original contour deforms to a sum of all three contours ($\mathbb{R} = C_{-1} + C_0 + C_1$), and all three saddle points contribute. So we have

$$Z = \boxed{\mathcal{J}_{-1}} + \mathcal{J}_0 + \boxed{\mathcal{J}_1}. \quad (3.18)$$

We have boxed both \mathcal{J}_{-1} and \mathcal{J}_1 because they are exactly degenerate and exponentially larger than \mathcal{J}_0 . If we perform the naive saddle-point approximation, we would obtain

$$Z \sim \mathcal{I}_{-1} + \mathcal{I}_0 + \mathcal{I}_1. \quad (3.19)$$

Now we see a confusion; since \mathcal{I}_0 is imaginary, it seems that Z might have an imaginary part in the approximation, even though it is a convergent real integral and thus is clearly actually real. What is happening?

As one can see in Fig. 8, if we really compute \mathcal{J}_1 exactly, integrating along C_1 , it will have a real part from the real-line part of the contour and an imaginary part from the imaginary-line part. The sum of the imaginary parts of \mathcal{J}_1 and \mathcal{J}_{-1} is exactly the negative of the imaginary part of \mathcal{J}_0 —they are simply integrating along the same contour in opposite directions. However, when we use the saddle-point approximation, \mathcal{I}_1 and \mathcal{I}_{-1} are both real; their imaginary parts are (rightfully) discarded as they are exponentially small. As we discussed before, in the saddle-point approximation, like Eq. (3.19), exponentially subdominant terms are meaningless. Thus, to be consistent, we should also drop \mathcal{I}_0 since \mathcal{J}_0 is exponentially suppressed, as indicated in Eq. (3.18). The integral is in fact real at any order in any expansion, if all the pieces of the same order are consistently kept.

Thus, although the saddle-point approximation can seem to produce an imaginary part in a real quantity, this is an illusion. Within a consistent expansion, real integrals are real. To get an imaginary part, we really do need to change the original contour of integration, as we explain next.

2. Unstable potentials and analytic continuation

To understand the imaginary part and factor of $\frac{1}{2}$, a standard example [5,6] is the following action function:

$$S_g(z) = \frac{z^2}{2} - g \frac{z^4}{4}. \quad (3.20)$$

Although this action is unbounded from below and unphysical, we will see in Sec. III B 3 how stabilizing it by adding a term $\frac{z^6}{60}$ to the action will lead to a similar result.

As above, we would like to study the imaginary part of

$$Z_g \equiv \int_{-\infty}^{\infty} dz e^{-\frac{1}{\hbar} S_g(z)} \quad (3.21)$$

for $g = 1$.

For any $g \neq 0$ the action has three saddle points: $s_j = \left\{ -\sqrt{\frac{1}{g}}, 0, \sqrt{\frac{1}{g}} \right\}$ for $j = -1, 0, 1$, respectively. The $z = s_0 = 0$ saddle point plays the role of the false-vacuum solution, and the $z = s_{\pm 1} = \pm \sqrt{\frac{1}{g}}$ saddle points play the role of the bounce. Applying the saddle approximation,

$$\begin{aligned} \mathcal{I}_{-1} &= \sqrt{-\pi\hbar} \exp\left(-\frac{1}{4g\hbar}\right) & \mathcal{I}_0 &= \sqrt{2\pi\hbar} \\ \mathcal{I}_1 &= \sqrt{-\pi\hbar} \exp\left(-\frac{1}{4g\hbar}\right), \end{aligned} \quad (3.22)$$

we find that \mathcal{I}_0 is real and $\mathcal{I}_{\pm 1}$ are imaginary.

As explained in the previous section, we cannot trust the saddle-point approximation because exponentially small imaginary contributions from the \mathcal{I}_0 have been dropped. Even worse in this case, Z itself does not converge at $z = \pm\infty$, so we cannot integrate Z along the real axis. However, Z_g can be integrated along the real axis for $g < 0$. The steepest-descent contours are shown for $g = 1$ and $g = -1$ in Fig. 9, with the black arcs indicating regions of convergence. So one thing we can do is calculate Z_g for $g < 0$ along the real axis and analytically continue that back to $g = 1$. We call that result Z_g^{cont} .

The convergent regions in the complex z plane of Z_g^{cont} change as a function of g . The analytic extension away from $g < 0$ is unique if we fix the end points of the contour to the convergent regions as we change g . For example, we can parametrize the analytic continuation from $g = -1$ to $g = 1$ by varying $g = -e^{i\theta}$ from $\theta = 0$ to $\theta = \pi$. For $\theta = 0$ the integration contour is given by the real line. As we change θ we can have the end points of the contour following the convergent region by rotating the contour of integration by an angle $\frac{\theta}{4}$ when we rotate $g = -1$ to $g = -e^{i\theta}$ (see Fig. 10). Note that rotating by $\theta = -\pi$ gives the conjugate result to rotating by $\theta = \pi$, and thus there is a branch cut along the positive real axis if we restrict the Riemann surface to a single sheet. In Fig. 9 the steepest descents are given for g slightly above and below the positive real axis to break the degeneracy of steepest-descent curves.

Fixing the end points of the contour, the rest of the contour can safely be deformed since $S_g(z)$ contains no poles. We can deform the $\theta = \pi$ contour of integration to the contour of steepest descent through the false-vacuum saddle point, C^+ . This can easily be seen by comparing Figs. 9 and 10. Similarly, rotating by $\theta = -\pi$ gives the contour C^- .

We therefore find

$$Z_{g=1}^{\text{cont}\pm} = \int_{C^\pm} dz \exp\left(-\frac{z^2}{2} + \frac{z^4}{4}\right) \quad (3.23)$$

where $(Z_g^{\text{cont}+})^* = Z_g^{\text{cont}-}$.

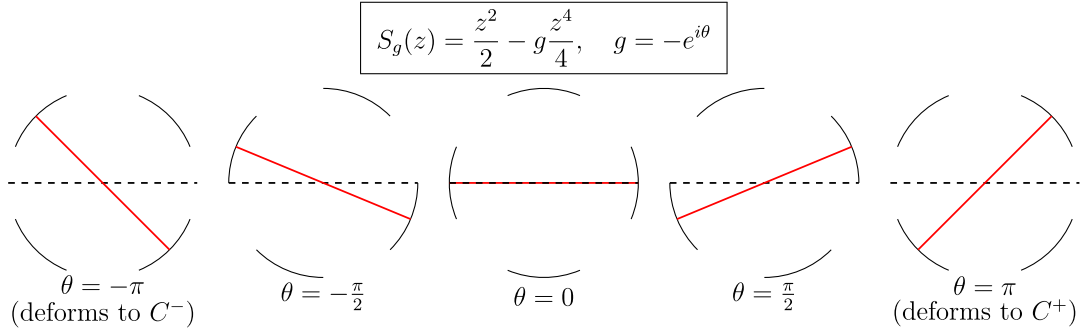


FIG. 10. The contour (red) used for analytic continuation as a function of θ compared to the original real axis contour (dashed line). Because there are no poles in $S(z)$, we can deform the contours however we like as long as they end in the same convergent regions [black arcs are where $\text{Re}(gz^4) < 0$].

Next, we want to approximate this integral using a saddle-point approximation. Unfortunately, the saddle-point approximation along the contour C^+ (or C^-) around the FV saddle point is purely real, to all orders in the expansion. The saddle-point approximation probes the function only in an asymptotically small region close to the saddle point. This is completely insensitive to the part of the contour C^+ (or C^-) which passes into the complex plane. However, it is precisely this part of the contour we are interested in, since that is where the imaginary part will come from. A way around this is to use the fact that $Z_g^{\text{cont}\pm}$ are complex conjugates of each other to write

$$\text{Im} Z_{g=1}^{\text{cont}+} = \frac{Z_{g=1}^{\text{cont}+} - Z_{g=1}^{\text{cont}-}}{2i} = \frac{\text{disc}(Z_{g=1}^{\text{cont}})}{2i}. \quad (3.24)$$

As can be seen in Fig. 11, integrating along $C^+ - C^-$ is equivalent to integrating along $C_{-1} + C_1$, that is, along the contours of steepest descent passing through the bounces.

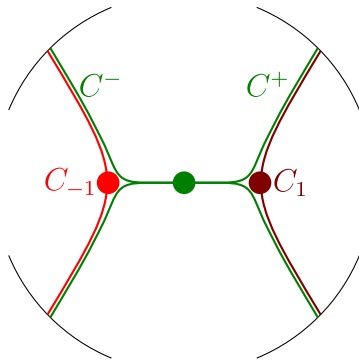


FIG. 11. The difference between the two FV contours (green lines, C^+ and C^-) is equivalent to the sum of the two bounce contours (C_{-1} and C_1). (Computing the discontinuity $\text{disc } Z_{g=1}^{\text{cont}}$ rather than just $Z_{g=1}^{\text{cont}}$ involves subtracting two integration contours, yielding the combined contour which can be deformed into a sum of the red and blue steepest-descent curves, with no presence of the green curve.

Now that we have directly related the imaginary part of $Z_{g=1}^{\text{cont}+}$ to an integral along paths which are complex at the saddle points, we can safely take the saddle-point approximation. Thus,

$$\text{disc } Z_{g=1}^{\text{cont}} = \mathcal{J}_{-1} + \mathcal{J}_1 \sim \mathcal{I}_{-1} + \mathcal{I}_1 \quad (3.25)$$

and therefore

$$\text{Im } Z_{g=1}^{\text{cont}+} \sim \frac{1}{2} \text{Im}(\mathcal{I}_{-1} + \mathcal{I}_1). \quad (3.26)$$

That is, the desired imaginary part is given by *half* the sum over the relevant bounces.

3. Bounded potentials and steepest-descent contours

One concern with the procedure described above is that the action we used in Sec. III B 2, $S_g[z] = \frac{1}{2}z^2 - g\frac{z^4}{4}$, was unbounded from below, and therefore unphysical. One might worry that the justification for analytic continuation and changing the integration contour from the real axis was to make the integral well defined; for a physical potential, perhaps the imaginary part remains zero. With that motivation, consider the action function

$$S_g(z) = \frac{z^2}{2} - g\frac{z^4}{4} + \frac{z^6}{60}. \quad (3.27)$$

This is similar to the action from Sec. III B 3, but now the integral Z over the real axis is convergent.

Now the action is sixth order, so $S'(z) = 0$ has five solutions for five saddle points. For $g = 1$, these are around $z \approx \{-3.0, -1.0, 0, 1.0, 3.0\}$, all along the real axis. There are six convergent regions, including the region at $z = \pm\infty$ for any g . So in this case, the original contour of the real line is perfectly fine for any g . Indeed, the function $Z_g = \int_{-\infty}^{\infty} dz \exp(-S_g(z))$ is an analytic function of g ; it is real for real g and has no discontinuity near $g = 1$. Thus, whatever we do, we are certainly not analytically continuing Z_g .

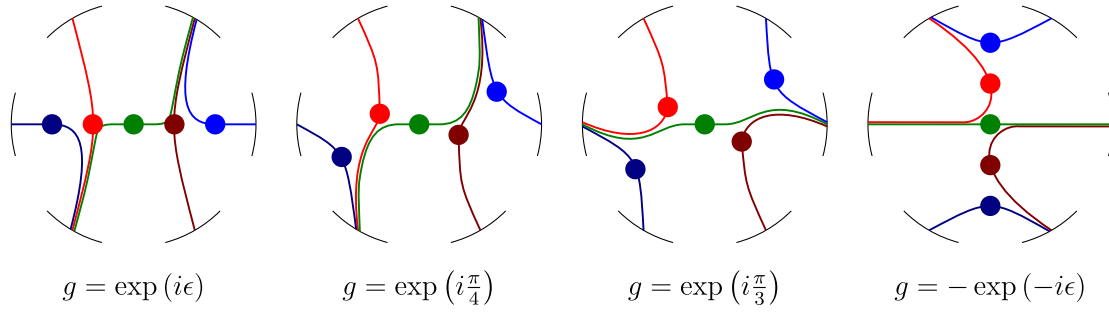


FIG. 12. For the action function $S_g[z] = \frac{z^2}{2} - g\frac{z^4}{4} + \frac{z^6}{60}$, the real axis is a valid integration contour for any g . However, as we change g , the steepest-descent contours change. We see that for $g = \exp(i\epsilon)$ the steepest-descent contour ends along the ray $\theta = \frac{\pi}{3}$ rather than the real axis, $\theta = 0$.

So what can we do? We saw in the previous section that for unbounded potentials, analytic continuation is equivalent to integrating along the steepest-descent contour through the false-vacuum saddle point. In the case of a bounded potential like we have here, analytic continuation and integrating along the contour of steepest descent are necessarily different; the steepest descent will always move off the real axis and will end in a different region of convergence. We therefore introduce the path integral along the steepest-descent contour through the FV saddle point,

$$Z^{C^\pm} = \int_{C^\pm} dz e^{-S_g(z)} \quad (3.28)$$

where C^\pm are the steepest-descent contours for $g = 1 \pm i\epsilon$. In Fig. 12 we see how the contours and saddle points move about as we rotate g .

By using Eq. (3.28) for both bounded and unbounded potentials, we will always find that the imaginary part of Z^{C^+} is given by

$$\text{Im } Z^{C^+} = \frac{1}{2i} \text{disc}(Z^{C^+}) \quad (3.29)$$

which in the saddle-point approximation is equal to $\frac{1}{2}$ times the sum over the relevant bounce saddle points. With this method, modifying the action far away from the region relevant to the false vacuum and the bounce does not seem to affect the prediction for the tunneling rate by very much, which is reassuring.

4. Dependence on the choice of contour

We have intimated that the key to finding the decay rate is to integrate not along the real axis but along a steepest-descent contour passing through the false vacuum. If the potential is deformed to be convex with the FV at the true minimum, the real axis will coincide with this contour. To clarify the importance of the FV saddle, let us now look at how different results arise when different saddle points are stabilized.

To explore how different steepest-descent contours may affect the result, consider the following two-parameter family of actions:

$$S_g(z) = h\frac{z}{12} - g\frac{z^2}{2} + \frac{z^4}{4}. \quad (3.30)$$

The case of interest is $g = h = 1$. There are three saddle points: a FV at $z \approx -1.0$, a bounce at $z \approx 0.0$, and a shot at $z \approx 1.0$. The saddle points and steepest-descent contours are shown in Fig. 13.

First, we consider keeping $h = 1$ and rotating to $g = -1$. For $h = 1$ and $g = -1$, the potential has one minimum at $z \approx 0.0$: The bounce has been stabilized. In this case, the integration contour along the z axis coincides with the bounce saddle-point contour. When we rotate back to $g = 1$, this contour lines up with the imaginary axis. Thus, integrating along the contour will give the complete imaginary part of the bounce saddle-point integration, *without* the factor of $\frac{1}{2}$. Of course, this had to happen: By stabilizing the bounce, we matched the integration contour with the bounce contour. When we rotate back, it remains lined up, and therefore the full integral over the bounce contour is kept.

Next, consider keeping $g = 1$ but rotating h from 1 to something negative and large enough to remove the other minimum, such as $h = -5$. For example, we can rotate as $h = -2 + 3e^{i\theta}$ with $0 \leq \theta \leq \pi$. For $h = -5$, the saddle point on the real axis is the FV saddle, and the other two have moved into the complex plane. When we rotate back to $h = 1$, this FV saddle moves along the real axis and then up *half* of the bounce saddle. Thus, for the h deformation, we get the extra factor of $\frac{1}{2}$, as expected.

Finally, consider rotating h from 1 to 5. This stabilizes the shot. Rotating back to $h = 1$, we see that the shot contour lines up with the other hand of the bounce contour when we stabilized the FV. Thus, we get a factor of $\frac{1}{2}$ in this case. The sign of the imaginary part in any case has to be fixed by physics. When we stabilize the

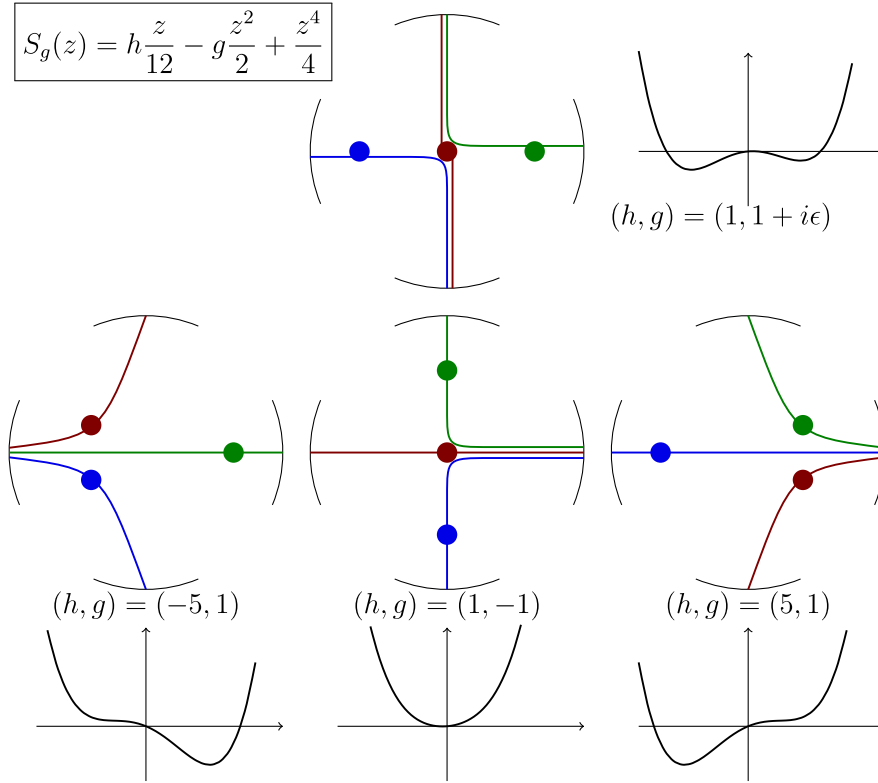


FIG. 13. The function $S_{h,g}(z) = h\frac{z}{12} - g\frac{z^2}{2} + \frac{z^4}{4}$ can be deformed in different ways from the physical case $(h, g) = (1, 1)$ (top). For $(h, g) = (-5, 1)$, the FV is stabilized, and the real axis lines up with the FV contour (green). The correct factor of $\frac{1}{2}$ results. For $(h, g) = (1, -1)$, the bounce is stabilized (red). The imaginary part computed in this way is the naive one, missing the factor of 2. For $(h, g) = (5, 1)$, the shot is stabilized (blue).

shot, we can choose the sign to be negative so that $\Gamma < 0$. This makes sense physically because flux enters the true vacuum region, so the probability grows with time. This corresponds to incoming Gamow boundary conditions (as in Sec. II B or Appendix A), and one expects $\Gamma < 0$.

C. Summary of potential deformation method

In this section, we discussed how to compute a decay rate from the Euclidean path integral, filling in some details and examining some peculiarities not mentioned in [1, 11] or elsewhere in the literature to our knowledge. In this method, one starts with a Euclidean path integral or partition function,

$$Z = \int_{x(0)=a}^{x(\mathcal{T})=a} \mathcal{D}x e^{-S_E[x]}, \quad (3.31)$$

which is real for all \mathcal{T} .

To get an imaginary part, we reduced the problem to integrating along a one-parameter family of curves, passing through the static FV path, the bounce, and the shot. Around the bounce saddle point $z = s_1$, the 1D integral is approximately

$$Z = \int dz e^{-S(z)} \approx \sqrt{\frac{2\pi}{S''(s_1)}} e^{-S(s_1)}, \quad (3.32)$$

which is imaginary because of the negative curvature around the bounce. However, one cannot just pick a single saddle point. One must use the proper integration contour, which can include multiple saddles, and the imaginary parts can (and do) cancel. Physically, the $\mathcal{T} \rightarrow \infty$ limit always picks out the true vacuum, and the path integral is dominated by the shot not the bounce. The path integral around the shot is real and exponentially larger than the path integral around the bounce. Instead, we want the metastable resonance state near the false vacuum to dominate.

To isolate the resonance near the FV, one approach is to deform the potential so that the FV is the true minimum. We write $g = 1$ for the original potential, and $g = -1$ for when the FV is stable. Unfortunately, at $g = -1$, the path integral is still real. Moreover, one cannot simply analytically continue the result back to $g = -1$, for then the true minimum would be the true vacuum again, and Z would still be real.

The right way to isolate the resonance is to deform the theory at $g = -1$ in a different way: We pin the contour

with the steepest-descent contour C^+ passing through the FV saddle. Near $g = -1$ this is identical to the analytic function Z_g . However, as the potential deforms back, the analytic continuation remains real, while pinning the contour to C^+ does not. This is a consequence of the fact that through the deformation, the system crosses a Stokes line, changing the region (the black arcs in Fig. 13, for example) in which the steepest descent through the FV saddle ends.

Of course, if at the end of the day all we want is to stick to the FV contour, we do not have to bother with the analytic continuation at all. The procedure is simply as follows:

- (i) The decay rate is given by the imaginary part of the path integral along the steepest-descent contour C^+ passing through the false-vacuum saddle point (*not* the bounce saddle point, and *not* the path integral over real paths).

Pinning the contour to focus on the FV is exactly what we want physically: We want the resonance associated with the FV to dominate, even when it is not the dominant saddle point. This procedure enforces this dominance, albeit in a somewhat artificial mathematical way.

The integral along the FV contour will produce something complex, $Z = \mathcal{R} + i\mathcal{I}$. Then the decay rate, in the limit $\mathcal{I} \ll \mathcal{R}$, is

$$\frac{\Gamma}{2} = \frac{1}{T} \text{Im} \ln Z \approx \frac{1}{T} \frac{\mathcal{I}}{\mathcal{R}}. \quad (3.33)$$

Unfortunately, to compute Z one cannot directly use the saddle-point approximation since close to the saddle point the FV contour will give something real, \mathcal{R} alone. This real part \mathcal{R} is indeed given by the saddle-point approximation around FV, $\mathcal{R} \approx \exp\{-S[\tilde{x}_{\text{FV}}]\}$. The imaginary part \mathcal{I} comes from a region on the contour far away from the FV saddle. Thus, to get the imaginary part under a saddle-point approximation, we have to compute the discontinuity between the two degenerate steepest-descent contours, which is equivalent to evaluating $\frac{1}{2}$ of the steepest-descent contour through the bounce saddles. That is,

- (i) the imaginary part of the path integral can be computed along the steepest-descent contour passing through the bounce saddle point, times $\frac{1}{2}$.

Thus, to leading order

$$\frac{\Gamma}{2} \sim \frac{1}{2T} \tilde{K} \frac{\exp\{-S[\tilde{x}_{\text{bounce}}]\}}{\exp\{-S[\tilde{x}_{\text{FV}}]\}}, \quad (3.34)$$

with the prefactor \tilde{K} computable systematically in the saddle-point approximation.

We have performed some helpful sanity checks on this algorithm. First, we considered examples from the literature that use actions unbounded from below, where one can get a nontrivial result using analytic continuation

because the integral Z along the real axis diverges. Thus, it looks like Z *must* be defined through analytic continuation. But when considering bounded potentials, we saw that analytic continuation does not give the right answer. Instead, pinning to the steepest-descent contour through the false vacuum gives consistent results in both the bounded and unbounded cases we have considered. Moreover, deforming the potential away from the FV region and the barrier has little effect on the rate, as desired.

Second, we checked that the deformation must stabilize the false vacuum, not any of the other saddle points. We found that if the bounce is stabilized, one gets an answer that is a factor of 2 too large. We also found that stabilizing the shot gives an answer with the wrong sign. These observations are consistent with the general physical argument that a proper derivation must include consideration that T not be too large. For very large T , only the true vacuum is relevant. By deforming the potential so the false vacuum is the ground state, the false-vacuum bound states can be found. After deforming back, these presumably turn into the resonances, with outgoing boundary conditions and imaginary energies.

There are two nagging questions we have been unable to answer in our explorations of this method.

1. How can the rate be calculated without using the saddle-point approximation, for example, nonperturbatively?

If the procedure were just analytic continuation, one could, in principle, compute the path integral as a function of g and analytically continue it from around $g = -1$ back to $g = 1$. However, fixing one dimension of an infinite-dimensional integral to a particular contour does not have an obvious nonperturbative analog.

2. Can one prove that pinning the contour to C^+ is identical to imposing the limits $T_{\text{slosh}} \ll T \ll T_{\text{NL}}$?

In other words, how do we know the number Γ computed through this mathematically consistent procedure always gives the decay rate exactly?

The next section discusses a more physical approach, with fewer mathematical subtleties, which we can (in principle) compare to the potential-deformation approach order by order.

IV. DIRECT METHOD

The potential-deformation method described in Sec. III connects the Euclidean action to the decay rate in a roundabout manner. It relies critically on an understanding of the subtleties of analytic continuation, steepest-descent contours, and saddle approximations of the path integral in order to obtain an imaginary number. In this section, we describe an alternative derivation that connects the path integral directly to the decay rate.³

³Another approach, employing coherent states, is described in [38].

Let us assume we have a potential with a FV, a barrier, and a true vacuum region (R), as in Fig. 1. The energy eigenstates which have support in the FV region are in bands of width Γ_i around resonance energies E_i . In Sec. II B, we were led to assume, for simplicity, that our initial state only had support for energies near E_0 . We would still like this to be true, but it is more convenient with path integrals to work with position eigenstates than energy eigenstates. So let us assume now that the initial state is localized at the point $x = a$ where the minimum of the well is located. As we take T large, this wave function will have dominant support from within the lowest energy band. It will also have some support for the higher energy bands; however, the higher energy components die off much faster than the E_0 components, so if we focus on time scales $T \gg E_0^{-1}$, we should be able to ignore those components.

The decay rate to a region R is defined as in Eq. (2.7):

$$\Gamma_R = \lim_{\substack{T/T_{\text{NL}} \rightarrow 0 \\ T/T_{\text{slosh}} \rightarrow \infty}} \frac{1}{P_{\text{FV}}(t)} \frac{dP_R(T)}{dT}, \quad (4.1)$$

where the probability P_R is defined as

$$P_R(t) = \int_{\text{R}} dx_f |\langle x_f; t | a; 0 \rangle|^2 = \int_{\text{R}} dx_f |\mathcal{N} D_F(a, 0; x_f, t)|^2. \quad (4.2)$$

The factor of \mathcal{N} in Eq. (4.2) is a normalization factor pulled out of the path-integral form of the propagator, so the Feynman propagator D_F is simply

$$D_F(a, 0; x_f, t) = \int_{x(0)=a}^{x(t)=x_f} \mathcal{D}x e^{iS[x]}. \quad (4.3)$$

Normally, there would be a factor of \mathcal{N} in Eq. (4.3), but we have put the \mathcal{N} in Eq. (4.2) instead.

Let us denote by b the point on the boundary of R where the potential is degenerate with the initial point a : $V(b) = V(a)$. By splitting every path into the part before it first hits b (at time t_0) and the part after t_0 , we can write

$$D_F(a, 0; x_f, t) = \int_0^t dt_0 \bar{D}_F(a, 0; b, t_0) D_F(b, 0; x_f, t - t_0) \quad (4.4)$$

where

$$\bar{D}_F(a, 0; b, t_0) \equiv \int_{x(0)=a}^{x(t_0)=b} \mathcal{D}x e^{iS[x]} \delta(t_b[x] - t_0). \quad (4.5)$$

Here, $t_b[x]$ is the functional returning the time the path $x(t)$ first crosses b . So \bar{D}_F is the Feynman propagator over paths on the interval $(0, t_0)$ that hit b exactly once, at $t = t_0$.

The separation in Eq. (4.4) works so long as all paths in the original propagator pass through b at least once. In the path integral, for each path $x(t)$ into R, there is a time t_0 when the path exits the barrier for the first time. Since we have taken b to be the classical turning point on the boundary of R, any path into R must hit b , so we can indeed use Eq. (4.4). Thus, we find

$$\begin{aligned} P_R(t) &= \mathcal{N} \mathcal{N}^* \int dx_f \int_0^T dt_0 \int_0^T dt'_0 \bar{D}_F(a, 0; b, t_0) \\ &\quad \times \bar{D}_F^*(a, 0; b, t'_0) D_F(b, 0; x_f, T - t_0) \\ &\quad \times D_F^*(b, 0; x_f, T - t'_0). \end{aligned} \quad (4.6)$$

Now we want to use the fact that once the particle gets to region R, it stays in region R; this is the limit $T/T_{\text{NL}} \rightarrow 0$. This fact lets us replace the sum of $|x_f\rangle\langle x_f|$ over points in R to a sum over all points. Such a replacement will modify P_R only by terms which are exponentially small, for example, suppressed by extra factors of the e^{-W} WKB penetration factor. Such exponentially small corrections are an irreducible ambiguity in what is meant by a decay rate. Thus, to the extent that Γ is well defined at all, we can drop them. Then,

$$\begin{aligned} &\int dx_f D_F(b, 0; x_f, T - t_0) D_F^*(b, 0; x_f, T - t'_0) \\ &= \int dx_f \langle b | e^{-iH(T-t_0)} | x_f \rangle \langle x_f | e^{iH(T-t'_0)} | b \rangle \\ &\approx D_F(b, 0; b, t'_0 - t_0). \end{aligned} \quad (4.7)$$

So we now have

$$\begin{aligned} P_R(t) &= \mathcal{N} \mathcal{N}^* \int_0^T dt_0 \int_0^T dt'_0 \bar{D}_F(a, 0; b, t_0) \\ &\quad \times \bar{D}_F^*(a, 0; b, t'_0) D_F(b, 0; b, t'_0 - t_0). \end{aligned} \quad (4.8)$$

Breaking the two integrals into the regions with $t'_0 > t_0$ and $t_0 > t'_0$, we can use Eq. (4.4) on the two halves along with $D_F^*(x, 0; y, t) = D_F(y, 0; x, -t)$ to get

$$\begin{aligned} P_R(t) &= \mathcal{N} \mathcal{N}^* \int_0^T dt_0 [\bar{D}_F(a, 0; b, t_0) D_F^*(a, 0; b, t_0) \\ &\quad + \bar{D}_F^*(a, 0; b, t_0) D_F(a, 0; b, t_0)]. \end{aligned} \quad (4.9)$$

Then expanding the definitions of D_F and \bar{D}_F in Eqs. (4.3) and (4.5) and plugging this into the definition of Γ_R in Eq. (4.1) produces

$$\Gamma_R = \frac{\mathcal{N}\mathcal{N}^*}{P_{\text{FV}}(T)} \left(\int_{x(0)=a}^{x(T)=b} \mathcal{D}x e^{iS[x]} \delta(t_b[x] - T) \right) \times \left(\int_{x(0)=a}^{x(T)=b} \mathcal{D}x e^{iS[x]} \right)^* + \text{c.c.} \quad (4.10)$$

Now let us consider how to calculate these two path integrals. Because the path integral has an imaginary exponent, it is not convergent (when integrated along the real x). For this reason the actual definition of the Minkowski path integral involves a strange imaginary integration path over x , or more simply, evaluating it for $T = iT$ real and then analytically continuing.

For the first path integral in Eq. (4.10), we analytically continue to $T = iT$. Then the boundary conditions are $x(0) = a$ and $x(T) = b$ as before, which are equivalent to $x(-T) = a$ and $x(0) = b$. For the second integral, which is complex conjugated, we must analytically continue to $T' = -i\tau = -T$ to ensure convergence. This leads to $x(0) = a$ and $x(-T) = b$ as boundary conditions, or equivalently, $x(0) = b$ and $x(T) = a$. Because the end points have switched, the two Euclidean path integrals can then be recombined, leading to

$$\Gamma_R = \frac{\mathcal{N}\mathcal{N}^*}{P_{\text{FV}}(T)} \int_{x(-T)=a}^{x(T)=a} \mathcal{D}x e^{-S_E[x]} [\eta_+ \delta(-i\tau_b[x]) + \eta_- \delta(i\tau_b[x])]. \quad (4.11)$$

To get to this result, we have replaced $t_b[x]$ with $-i\tau_b[x]$ in the first δ function and $i\tau_b[x]$ in the second δ function. However, changing variables within a δ function has to be done carefully since the δ function is not defined for imaginary arguments. Thus, we have added phase factors η_{\pm} in front, which we will now fix.

In order to analytically continue correctly, we should first remove the δ -function mode, then analytically continue, and then put it back. This will lead to $\delta(\pm if[x]) \equiv \eta_{\pm} \delta(f[x])$ for some η_{\pm} . To see what these η_{\pm} are, we focus only on a single degree of freedom restricted by the δ function. We analytically continue as follows:

$$\begin{aligned} & \int dx [A e^{iaS(x)} \delta(\alpha t[x]) + A^* e^{-iaS(x)} \delta(\alpha t[x])] \\ &= \left(\int dx [\eta_- A e^{-rS(x)} \delta(irt[x]) + \eta_+ A^* e^{-sS(x)} \delta(-ist[x])] \right)_{\substack{r=-i\alpha \\ s=i\alpha}}, \end{aligned} \quad (4.12)$$

where A represents the integral over all the other modes, and α represents a phase-factor coefficient in front of τ which we will rotate from 1 to i . This notation means that we evaluate the integrals for r and s positive (where the other path integrals converge) and then make the substitutions. Integrating over the δ functions gives

$$\begin{aligned} & \frac{A e^{iaS(x_*)} + A^* e^{-iaS(x_*)}}{|\alpha t'(x_*)|} \\ &= \left(\frac{\eta_- A e^{-rS(x_*)}}{r|t'(x_*)|} + \frac{\eta_+ A^* e^{-sS(x_*)}}{s|t'(x_*)|} \right)_{\substack{r=-i\alpha \\ s=i\alpha}} \end{aligned} \quad (4.13)$$

$$= \frac{1}{r|t'(x_*)|} (\eta_- A e^{-rS(x_*)} - \eta_+ A^* e^{rS(x_*)})_{r=-i\alpha} \quad (4.14)$$

from which we see that $\eta_+ = -\eta_- = \text{isign}(\alpha)$ will give us the right answer. Thus, we see that the analytic continuation of the δ function is

$$\begin{aligned} & \int dx [A e^{iaS(x)} \delta(\alpha t[x]) + A^* e^{-iaS(x)} \delta(\alpha t[x])] \\ &= -\text{sign}(\alpha) 2\text{Im} \left(A \int dx e^{-rS(x)} \delta(rt(x)) \right)_{\substack{r>0 \\ r=-i\alpha}}, \end{aligned} \quad (4.15)$$

where the notation means that we evaluate the expression for $r > 0$, and then afterwards analytically continue to $r = -i\alpha$. The sign ambiguity is due to a branch cut: The answer depends on which way we rotate the argument, and we will have to fix it with a physical argument.

Fixing the sign by requiring $\Gamma > 0$, the precise version of Eq. (4.11) is then

$$\Gamma = \lim_{T \rightarrow \infty} \left| \frac{\mathcal{N}\mathcal{N}^*}{P_{\text{FV}}(T)} 2\text{Im} \left(\int_{x(-T)=a}^{x(T)=a} \mathcal{D}x e^{-S_E[x]} \delta(\tau_b[x]) \right) \right|_{\substack{T>0 \\ T=iT}}. \quad (4.16)$$

The normalization $\frac{P_{\text{FV}}(T)}{\mathcal{N}\mathcal{N}^*}$ can be manipulated in a similar manner in the limit $T/T_{\text{NL}} \rightarrow 0$, which ultimately gives us

$$\Gamma = \lim_{T \rightarrow \infty} \left| 2\text{Im} \left(\frac{\int \mathcal{D}x e^{-S_E[x]} \delta(\tau_b[x])}{\int \mathcal{D}x e^{-S_E[x]}} \right) \right|_{\substack{T>0 \\ T=iT}} \quad (4.17)$$

where both path integrals are evaluated with boundary conditions $x(\pm T) = a$. This formula provides an exact expression for the decay rate defined in Eq. (2.7).

In Sec. III, in the potential-deformation method, we were worried that we might accidentally be taking the imaginary part of a real quantity by making an invalid saddle-point approximation. So let us now discuss in detail where the imaginary parts are coming from in this direct method. Equation (4.10), before the analytic continuation to imaginary time, is all-orders exact and manifestly real; one could, in principle, make a lattice and calculate it numerically. Then we analytically continued Eq. (4.10) to arrive at Eq. (4.17). The path integral $\int \mathcal{D}x e^{-S_E}$, without the δ function, would be real for real T , but the δ function will introduce a factor with dimensions of time, which becomes imaginary when we plug in $T \rightarrow iT$. Thus, we are taking the imaginary part of something purely imaginary in

Eq. (4.17). We will discuss the saddle-point approximation's interaction with this story in Secs. IV A and IV B.

Before showing how Eq. (4.17) can be evaluated, let us contrast it with the potential-deformation method discussed in Sec. III. To make a precise connection to Eq. (3.4), let us first change from \mathcal{T} to $\mathcal{T}/2$ (since the time is going to infinity, the factor of 2 has no effect). To match the other formula, we need to reintroduce the time-translation degeneracy. Isolating the term

$$\mathcal{E} \equiv \frac{\int \mathcal{D}x e^{-S_E[x]} \delta(\tau_b[x])}{\int \mathcal{D}x e^{-S_E[x]}} \quad (4.18)$$

which is time-translation invariant for all τ , that is,

$$\mathcal{E} = \mathcal{E}_\tau \equiv \frac{\int \mathcal{D}x e^{-S_E[x]} \delta(\tau_b[x] - \tau)}{\int \mathcal{D}x e^{-S_E[x]}}, \quad (4.19)$$

allows us to rewrite it as

$$\mathcal{E} = \frac{1}{\mathcal{T}} \int_{-\mathcal{T}/2}^{\mathcal{T}/2} d\tau \mathcal{E}_\tau = \frac{1}{\mathcal{T}} \frac{\int_{\text{paths hit } b} \mathcal{D}x e^{-S_E[x]}}{\int \mathcal{D}x e^{-S_E[x]}} \quad (4.20)$$

from which we arrive at

$$\Gamma = \lim_{\mathcal{T} \rightarrow \infty} \left| 2\text{Im} \left(\frac{1}{\mathcal{T}} \frac{\int_{\text{paths hit } b} \mathcal{D}x e^{-S_E[x]}}{\int \mathcal{D}x e^{-S_E[x]}} \right) \right|_{\substack{\mathcal{T} > 0 \\ \mathcal{T} = i\mathcal{T}}}. \quad (4.21)$$

The δ function in the numerator has been removed by the $\int d\tau$, except that it leaves the requirement that the path must hit b at some time so that τ_b is defined. Thus, the path integral in the numerator will exclude the constant false-vacuum solution which dominates the denominator. In this way, the need to determine the contour of steepest descent as we did in the potential-deformation method is side-stepped completely.

A. Saddle-point approximations

As discussed in Sec. IV, $\int \mathcal{D}x e^{-S_E}$, computed to all orders, is real. And as we saw in Sec. III B, when we approximate the path integral with a sum over saddle points, some of the saddle points might be imaginary. The imaginary parts will cancel if all the saddle points are kept associated with the integration contour, but if some can be dropped, the result may be complex. In the traditional method, deformation of the contour of integration in the path integral is used to justify dropping some saddle points, giving a well-defined imaginary part. In the direct method, the imaginary part comes out with less gymnastics. The path integral is real for real \mathcal{T} but simply becomes imaginary for imaginary \mathcal{T} .

In this section we will show that when performing the saddle-point approximation for real \mathcal{T} , the true vacuum

solution (the shot) dominates, but when evaluated for imaginary \mathcal{T} , the instanton solution dominates. Thus, we are justified in using only the instanton because we are looking at imaginary \mathcal{T} . In particular, there is no tension with the instanton's saddle-point expansion (which matters for imaginary \mathcal{T}) being imaginary when the original path integral is real (for real \mathcal{T}).

As in the potential deformation method, the path integrals in Eq. (4.17) are approximated by a sum of saddle points:

$$\frac{\exp(-S_{\text{shot}}) + \exp(-S_{\text{bounce}})}{\exp(-S_{\text{shot}}) + \exp(-S_{\text{bounce}}) + \exp(-S_{\text{FV}})}. \quad (4.22)$$

Consider first the denominator. It contains contributions from exactly the same paths as in the potential-deformation method, shown in Fig. 5: the static FV solution, the bounce, and the shot. Because of the forms of these solutions, it is clear that the \mathcal{T} dependence, for large \mathcal{T} , must have a linear dependence for the long stationary times and a constant piece for the brief times when the particle is rolling fast:

$$S_{\text{shot}} = E_{\text{TV}}\mathcal{T} + S_S^0, \quad (4.23)$$

$$S_{\text{FV}} = E_{\text{FV}}\mathcal{T}, \quad (4.24)$$

$$S_{\text{bounce}} = E_{\text{FV}}\mathcal{T} + S_B^0. \quad (4.25)$$

We also note that $S_S^0 > S_B^0$ since the shot must go faster than the bounce and hence has more energy.

Recall that in the potential-deformation method, the shot dominated for the actual path integral with the physical potential, but when we deformed to $g < 0$, then the false vacuum dominated. With the direct method, rather than deforming the potential, we perform the standard $\mathcal{T} \rightarrow i\mathcal{T}$ Wick rotation. For real \mathcal{T} , the shot dominates. But we are not interested in which dominates for real \mathcal{T} , but rather which dominates for $\mathcal{T} \rightarrow i\mathcal{T}$. Then,

$$S_{\text{shot}} = iE_{\text{TV}}\mathcal{T} + S_S^0, \quad (4.26)$$

$$S_{\text{FV}} = iE_{\text{FV}}\mathcal{T}, \quad (4.27)$$

$$S_{\text{bounce}} = iE_{\text{FV}}\mathcal{T} + S_B^0. \quad (4.28)$$

Since $S_B^0 < S_S^0$, due to the e^{-S} factors in the saddle-point approximation, the bounce exponentially dominates over the shot. However, both of these are dominated by the FV solution, which has no exponential suppression at all. Thus, for the denominator, if we drop exponentially suppressed pieces, only the FV contribution remains.

The numerator of Eq. (4.22) is similar to the denominator but has been modified by the $\delta(\tau_b)$. In particular, the FV solution, which never hits the point b , is removed entirely by the δ function. The shot is also removed since it hits b

before $\tau = 0$ (it hits the TV region at $\tau = 0$), but there is a solution qualitatively similar to the shot that we call the modified shot, or shot' as in Eq. (4.22).⁴ In any case, the argument for the numerator is then exactly the same as for the denominator; for real \mathcal{T} the modified shot dominates, but when we rotate $\mathcal{T} \rightarrow iT$, the constant part of the action now controls the size of e^{-S} and so the bounce dominates. Since the false vacuum is not present in the numerator at all, the result is given by the bounce alone.

In summary, performing the saddle-point approximation to Eq. (4.17) for imaginary \mathcal{T} carefully, we find that the bounce dominates the numerator and the FV dominates the denominator. For real \mathcal{T} , this would *not* be the correct set of saddle points to use (the correct saddle points would be the shot and modified shot). The point is that there is no tension between the dominant saddle points being imaginary (for imaginary \mathcal{T}) and the full path integral being real (for real \mathcal{T}).

B. Saddle-point approximation and NLO formula

Having the all-orders formula, Eq. (4.17), we want to apply the saddle-point approximation to it to get something we can actually calculate. If \bar{x} is the bounce solution to the Euclidean equations of motion and x_{FV} is the static solution which stays at the false vacuum, then we see that at leading order,

$$\Gamma_{\text{R}}^{\text{LO}} = \# \frac{e^{-S_E[\bar{x}]}}{e^{-S_E[x_{\text{FV}}]}} \quad (4.29)$$

which is the usual leading-order formula [40]. It also agrees with Eq. (2.18), since for $E_0 = 0$ in the false vacuum, $S_E = 2 \int \sqrt{2m\bar{V}} dx$.

To go to NLO, we would like to perform a Gaussian approximation on Eq. (4.17). Expanding around the bounce we write $x = \bar{x} + \delta x$ and

$$\Gamma = \frac{e^{-S_E[\bar{x}]}}{e^{-S_E[x_{\text{FV}}]}} \lim_{\tau \rightarrow \infty} \left| \frac{2\text{Im} \int [\mathcal{D}\delta x] e^{-\frac{1}{2}S_E''[\bar{x}]\delta x^2} \delta(\tau_b[\bar{x} + \delta x])}{\int [\mathcal{D}\delta x] e^{-\frac{1}{2}S_E''[x_{\text{FV}}]\delta x^2}} \right|. \quad (4.30)$$

Normally, for a Gaussian integral, we expand the path in orthonormal modes

⁴There will nevertheless still be a lower-action solution hitting b at $\tau = 0$ (we know this because the bounce still has a negative eigenvalue [1,39]). The minimum action solution probably looks like the bounce up to $\tau = 0$ spliced to a rescaled shot for $\tau > 0$. The shot part has to be rescaled to return to the FV at $\tau = \mathcal{T}$. The extra kick needed to splice these solutions at $\tau = 0$ is allowed because the δ function can introduce discontinuities in $\partial_\tau x(\tau)$. We call the actual minimum action solution the modified shot.

$$x^{\xi_0, \xi_1, \dots}(\tau) = \bar{x}(\tau) + \sum_{i=0}^{\infty} \xi_i x_i(\tau) \quad (4.31)$$

where $x_i(\tau)$ are the eigenvectors of $S_E''[\bar{x}]$ with eigenvalues λ_i . Plugging this back into Eq. (4.30) would then give Gaussian integrals $\int d\xi_i e^{-\frac{1}{2}\lambda_i \xi_i^2}$. However, since one of the modes has a zero eigenvalue [$\lambda_0 = 0$ for $\xi_i(\tau) \propto \partial_\tau \bar{x}$], this does not quite work. To resolve the divergent zero-mode integral, we must replace ξ_0 with a collective coordinate τ_0 [11,41–44] (see also [3,6,45]). This means that instead of Eq. (4.31), we parametrize our paths as

$$x^{\tau_0, \xi_1, \dots}(\tau) = \bar{x}(\tau - \tau_0) + \sum_{i=1}^{\infty} \xi_i x_i(\tau - \tau_0). \quad (4.32)$$

In the potential-deformation method, the integral over τ_0 gives a factor of \mathcal{T} (due to the exact translation symmetry) that resolves the $\int d\xi_0$ singularity, along with a Jacobian factor from going between ξ_0 and τ_0 ; one then divides by \mathcal{T} to find the rate. In the direct method, we can remove the collective coordinate with the δ function. With the parametrization in Eq. (4.32), the functional τ_b is

$$\tau_b[x^{\tau_0, \xi_1, \dots}] = \tau_0 + \tau_b[x^{0, \xi_1, \dots}]. \quad (4.33)$$

So depending on the $\{\xi_i\}$, one of two things happens:

1. If the path $x^{0, \xi_1, \dots}$ hits b at some time, then $\delta(\tau_b[x])$ simply removes the τ_0 integral and fixes τ_0 to some value $\tau_*(\xi_i)$.
2. If the path $x^{0, \xi_1, \dots}$ never hits b , then the δ function is always 0, and this point in ζ space does not contribute at all.

So we get

$$\Gamma^{\text{NLO}} = \frac{e^{-S_E[\bar{x}]}}{e^{-S_E[x_{\text{FV}}]}} \lim_{\mathcal{T} \rightarrow \infty} \left| \frac{2\text{Im} \int d^n \zeta \Theta[\zeta_i x_i(0)] J[\tau_*(\zeta), \zeta] e^{-\frac{1}{2} \sum \lambda_i \zeta_i^2}}{\int \mathcal{D}\delta x e^{-\frac{1}{2} S_E''[x_{\text{FV}}] \delta x^2}} \right| \quad (4.34)$$

where the $d^n \zeta$ indicates infinitely many integrals. We will now explain the addition of the theta function and the Jacobian factor, and we will see how an appealing feature of this method is the explanation of the factor of $\frac{1}{2}$.

Since the path \bar{x} just barely hits b at its maximum, the constraint that $x = \bar{x} + \delta x$ must hit b forces $\delta x(0) \geq 0$. Since $\delta x = \sum_i \xi_i x_i$, we can enforce this positivity constraint with a step function $\Theta[\zeta_i x_i(0)]$. Now, since we are working at Gaussian order only and this is a constraint on a simple linear combination of the ζ , we can use symmetry of the other terms under $\zeta \rightarrow -\zeta$ to drop the step function and divide by 2. This factor of 2, which arises in the Euclidean approach from a subtle analytic continuation argument

(cf. Sec. III B), arises naturally in the direct method from the requirement that the δ function fire. More physically, it is a requirement that the path enter the destination region DV, which excludes exactly half the variations around \bar{x} .

Finally, we must discuss the Jacobian $J(\tau_0, \zeta)$ arising when one goes from the orthonormal basis of fluctuations in Eq. (4.31) to the collective coordinate parametrization in Eq. (4.32). Note that J is nonsingular after fixing τ_0 , and it has some expansion in ζ . At NLO, we only need to keep the constant, ζ -independent piece. So we can replace

$$J(\tau_*(\zeta), \zeta) \rightarrow J(\tau_*(0), 0) = J(0, 0). \quad (4.35)$$

This Jacobian at leading order is well known [1,3,6,45] and discussed further in Appendix B⁵:

$$J(0, 0) = \sqrt{S_E(\bar{x})/m}. \quad (4.36)$$

Putting together the Jacobian factor and the factor of $\frac{1}{2}$, we get

$$\Gamma^{\text{NLO}} = \frac{e^{-S_E[\bar{x}]}}{e^{-S_E[x_{\text{FV}}]}} \sqrt{S_E[\bar{x}]/m} \left| \frac{1}{\sqrt{\pi}} \text{Im} \left(\frac{\det' \frac{1}{2} S_E''[\bar{x}]}{\det \frac{1}{2} S_E''[x_{\text{FV}}]} \right)^{-1/2} \right| \quad (4.37)$$

where \det' indicates the determinant omitting the 0 eigenvalue, and the boundary conditions of the determinants' domains are $x(\pm\infty) = a$. The π arises because the denominator path integral has one more Gaussian integral than the numerator.

While the dimensions of Eq. (4.37) are the correct dimensions of the rate, they have become obscured by the combination of the $\sqrt{S_E/m}$ and the determinants. To make the units clearer, let us remove $m/2$ from the determinant, using $\frac{\det' A}{\det A} \sim \frac{1}{A}$:

$$\Gamma^{\text{NLO}} = \frac{e^{-S_E[\bar{x}]}}{e^{-S_E[x_{\text{FV}}]}} \sqrt{S_E[\bar{x}]/2\pi} \left| \frac{\det' S_E''[\bar{x}]/m}{\det S_E''[x_{\text{FV}}]/m} \right|^{-1/2}. \quad (4.38)$$

Expanding S'' then gives

$$\Gamma^{\text{NLO}} = \frac{e^{-S_E[\bar{x}]}}{e^{-S_E[x_{\text{FV}}]}} \sqrt{\frac{S_E[\bar{x}]}{2\pi}} \left| \frac{\det' (-\partial_t^2 + \frac{V''(\bar{x}(t))}{m})}{\det (-\partial_t^2 + \frac{V''(a)}{m})} \right|^{-1/2}. \quad (4.39)$$

This agrees exactly with the formula surmised from the potential-deformation method [3,6].

⁵In the existing literature (e.g. [3]), authors often calculate $J(\tau_0 = 0)$, which is all that we need for our derivation. However, for *their* derivations using the potential-deformation method, they need the stronger derivation of $J(\tau_0)$ for general τ_0 . For this reason, in Appendix B, we prove that J is a constant function of τ_0 , even though in our case we could simply ignore the τ dependence.

C. Direct method in $d > 1$

In more than 1 dimension, the main change is that we must extend the turning point b to a surface Σ of possible turning points since paths can enter the destination region from any direction. The critical equation (4.4) becomes, in multiple dimensions,

$$D_F(a, 0; x_f, t) = \int_{\Sigma} db \int_0^t dt_0 \bar{D}_F(a, 0; b, t_0) D_F(b, 0; x_f, t - t_0) \quad (4.40)$$

for Σ any codimension-1 surface which all paths go through. The only subtlety is that, to avoid overcounting of paths that enter and leave, the functional $t_b[x]$ in \bar{D}_F only returns the first time $x(t)$ hits b if that is the first time the path crosses Σ at all (and returns ∞ otherwise).

From there the steps are the same as the one-dimensional case. Equations (4.6)–(4.8) contain two integrals $\int_{\Sigma} db \int_{\Sigma} db'$. Equation (4.10) will thus include an integral $\int_{\Sigma} db$, which stays through the end. Thus, we see

$$\Gamma_R = \left| \frac{\mathcal{N}\mathcal{N}^*}{P_{\text{FV}}(\infty)} 2\text{Im} \int_{\Sigma} db \int_{x(-\infty)=a}^{x(\infty)=a} \mathcal{D}x e^{-S_E[x]} \delta(\tau_b[x]) \right|, \quad (4.41)$$

where $\int_{\Sigma} db \delta(\tau_b[x]) = \delta(t_{\Sigma}[x])$, and t_{Σ} is the operator which returns the first time $\bar{x}(t)$ crosses Σ . Thus,

$$\Gamma_R = \left| \frac{2\text{Im} \int \mathcal{D}x e^{-S_E[x]} \delta(\tau_{\Sigma}[x])}{\int \mathcal{D}x e^{-S_E[x]}} \right| \quad (4.42)$$

where now Σ is the entire surface which bounds R, just like b was the turning point at the boundary of R. Both path integrals go from $x(-\infty) = a$ to $x(\infty) = a$.

D. Comparisons of the potential-deformation and direct approaches

While similar in many ways, the derivations in Secs. III and IV have a few key differences.

1. The potential-deformation method starts from $\frac{\Gamma}{2} = |\text{Im} E_0|$. While the decay rate is the imaginary part of an energy, as explained in Sec. II B, it is certainly not the imaginary part of the ground-state energy E_0 . There is an implicit assumption that deforming the potential (or, more honestly, integrating over complex paths intersecting the FV saddle) somehow isolates the energy of interest. The direct method instead starts from a physical definition, in Eq. (2.7), and there is no leap of faith required.
2. In both approaches, the path integral only has an imaginary part after an analytic continuation or

contour deformation. In the direct approach, this is the usual $\mathcal{T} \rightarrow iT$ Wick rotation, which came naturally in the derivation. In the Euclidean approach, this contour deformation had to be put in by hand; instead of just taking the imaginary part of E_0 , we had to deform the contour of integration in the path integral, in order to get a nonzero imaginary part.

3. Calculationally, the formulas remove the divergent Gaussian integral of the time-translation zero mode in the path integral differently. The deformation approach schematically generates $\Gamma\mathcal{T} = \int \mathcal{D}x e^{-S_E}$. The divergence at large \mathcal{T} is removed by dividing by \mathcal{T} . In the direct approach, $\Gamma = \int \mathcal{D}x e^{-S_E} \delta(\tau_b)$, and the would-be divergent integral is removed by the δ function.
4. In both methods, the NLO rate is given by a path integral around a bounce configuration divided by a path integral around the static FV solution. In the direct method, the FV solution does not contribute to the path integral in the numerator because it never fires the $\delta(\tau_b)$. In the deformation approach, to prevent the FV from dominating, one needs to calculate the discontinuity between the two degenerate steepest-descent contours.
5. In the deformation formula, the factor of $\frac{1}{2}$ in the NLO approximation arises because an integral along the FV contour has half of the imaginary part of an integral along the bounce contour. In the direct formula, it arises from the fact that only half of all small variations of the bounce solution enter the destination region R.
6. In the direct approach, one never has to worry about summing over approximate instantons or the validity of the dilute gas approximation. One simply systematically calculates the expansion of a path integral in \hbar .

V. TUNNELING IN QUANTUM FIELD THEORY

Quantum field theory is just quantum mechanics, where the Hilbert space happens to be an infinite-dimensional Fock space. Thus, all the fundamental facts of quantum mechanics apply, including all the facts about tunneling. Understanding tunneling in QFT is exactly the same as understanding tunneling in quantum mechanics (QM). However, some confusing language can make it more obscure than it needs to be.

Let us begin with theories with a single scalar field ϕ with a (classical) action of the form

$$\begin{aligned} S[\phi] &= \int d^4x \left[\frac{1}{2} (\partial_\mu \phi)^2 - V(\phi) \right] \\ &= \int d^3x dt \left[\frac{1}{2} \dot{\phi}^2 - \frac{1}{2} (\vec{\nabla} \phi)^2 - V(\phi) \right] \end{aligned} \quad (5.1)$$

where $\dot{\phi} \equiv \partial_t \phi$. We assume $V(\phi)$ has a false vacuum at $\phi = \phi_F$, where the potential is $V(\phi_F) = V_F$. It is

particularly convenient to shift the overall potential so that $V_F = 0$ since this is the unique value for which the total potential energy of the false vacuum $\int d^3x V_F$ is finite. Thus, when convenient, we will follow the usual convention and assume $\phi_F = 0$ and $V_F = 0$.

When going from finite-dimensional QM to infinite-dimensional QFT, it is important to keep in mind that the notation and language change as well:

1. The set of classical configurations is not an n -dimensional set of values \vec{x} , but the infinite-dimensional space of all field configurations, $\phi(x)$.
2. A quantum state, such as the true ground state, is not specified by a wave function, $\psi(x)$, but by a wave functional $\Psi[\phi]$.

In particular, in 1D quantum mechanics, we argued in Sec. II B that the decay rate is independent of the initial wave function, as long as it has some support of the lowest resonance in the false-vacuum region. Thus, for the path-integral derivation, we took it to be $\psi(x) = \delta(x - a)$. We could equally well have taken it to be the ground state of a harmonic oscillator in the quadratic approximation to the false-vacuum well: $\psi(x) \propto \exp(-\frac{1}{2} m \omega_a x^2)$ where $\omega_a = V''(a)$. Or we could have included admixtures of higher excited states or continuum modes if the domain of x was unbounded.

In quantum field theory, the ground state in the false vacuum is specified by a functional

$$\Psi[\phi] = \mathcal{N} \exp \left\{ -\frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \omega_k \tilde{\phi}(k) \tilde{\phi}(-k) \right\} \quad (5.2)$$

where $\omega_k = \sqrt{k^2 + m^2}$ and $\tilde{\phi}(k) = \int d^3x e^{ikx} (\phi(x) - \phi_F)$ are the Fourier modes of $\phi(x)$. States with a finite number of particles are finite-energy excitations on top of this vacuum. The wave functional corresponding to the first excited state of the harmonic oscillator has an infinite amount of energy (proportional to volume) more than the ground state. For the purpose of the tunneling rate calculation, we can, analogously to QM, take $\Psi[\phi] = \delta[\phi - \phi_F]$. This is what we mean when we say the state is initially localized at ϕ_F .

Another important difference is as follows:

3. Despite the fact that the Lagrangian has an object called V , the energy of a static field configuration $\phi(\vec{x})$ is not given just by $V[\phi]$ but it also includes a gradient energy contribution. Thus, we define the potential-energy functional as

$$U[\phi(x)] \equiv \int d^3x \left[\frac{1}{2} (\vec{\nabla} \phi)^2 + V(\phi) \right]. \quad (5.3)$$

The fact that the intuitive “potential” object is not the full potential energy through which the tunneling occurs leads to some effects that seem counterintuitive at first. For

example, suppose $V(\phi) = V_F$ is constant, or has a slight downward slope $V(\phi) = V_F - \varepsilon\phi$. Then one might imagine, since there is no barrier in V , that tunneling would occur infinitely fast, dominating over the classical rolling. However, there is a potential barrier in $U[\phi]$ due to the gradient energy term, so the tunneling rate is finite even as $\varepsilon \rightarrow 0$. See [46] for more details. We discuss below the case $V(\phi) = -\lambda\phi^4$ where one also might worry that tunneling would be infinitely fast.

In quantum field theory, the decay rate can be calculated to all orders using Eq. (4.42) with an infinite number of degrees of freedom:

$$\Gamma_R = \lim_{\tau \rightarrow \infty} \left| \frac{2\text{Im} \int \mathcal{D}\phi e^{-S_E[\phi]} \delta(\tau_\Sigma[\phi])}{\int \mathcal{D}\phi e^{-S_E[\phi]}} \right|. \quad (5.4)$$

Now Σ is a codimension-1 surface in the enormous configuration space of possible classical field configurations $\phi(\vec{x})$ which bounds the destination region R (also a region in field configuration space). The surface Σ is naturally taken to be the set of field configurations for which $U[\phi] = U[\phi_{FV}]$. Indeed, because semiclassical methods conserve energy, the end point of quantum tunneling must be on this surface.

Some interesting examples of the competition between tunneling rates to different regions were considered by Brown and Dahlen [47]. In particular, these authors explored examples where the rate for one of the tunneling processes abruptly goes from a finite value to zero after a small change in the potential. They also pointed out that in quantum field theory, for the surface Σ on which $U[\phi] = U[\phi_{FV}]$ is connected, one can interpolate between one region R and another with nonspherically symmetric configurations comprising subcritical and supercritical bubbles. Nevertheless, after tunneling occurs, subcritical bubbles will implode. In field theory, there is a well-defined set of quasistable local minima to which fields will evolve classically after tunneling occurs. Possible regions R are therefore naturally taken to be bounded by the separate parts of Σ from which fields will evolve separately to each minima.

In Sec. VB we will look at how we can visualize the decay in terms of the potential-energy functional along the particular field directions.

A. Bounces in QFT

If we want to work in the saddle-point approximation, then we need to find the stationary configurations of the Euclidean action:

$$\begin{aligned} S_E[\phi] &= \int d^4x \left[\frac{1}{2} (\partial_\mu \phi)^2 + V(\phi) \right] \\ &= \int d^3x dt \frac{1}{2} \dot{\phi}^2 + \int dt U[\phi]. \end{aligned} \quad (5.5)$$

The dominant bounce will always be $O(4)$ symmetric [48]; that is, $\phi_B(x) = \phi_B(\rho)$, where $\rho = \sqrt{x^2 + y^2 + z^2 + \tau^2}$. For these solutions, the Euclidean equations of motion reduce to

$$\partial_\rho^2 \phi_B + \frac{3}{\rho} \partial_\rho \phi_B - V'[\phi_B] = 0. \quad (5.6)$$

This equation has the analog mechanics interpretation of a particle rolling down a potential $-V(\phi)$ with ρ a time coordinate and $\frac{3}{\rho} \dot{\phi}$ representing a kind of time-dependent friction. Matching onto the false vacuum with $\phi = \phi_F$ at $\tau = \pm\infty$ translates to the boundary condition $\phi(\rho = \infty) = \phi_F = 0$. For the solution to be smooth at $\rho = 0$, we must also have $\partial_\rho \phi|_{\rho=0} = 0$. With these boundary conditions, the analog classical system has a particle starting at a point ϕ_0 at rest (at $\rho = 0$) and ending at rest at $\phi = \phi_F = 0$ at $\rho = \infty$.

One should think of $\phi_B(\vec{x}, \tau)$ as providing a path through field space $\phi(\vec{x})$ parametrized by τ . This path goes from the false vacuum at $\tau = -\infty$ to the bubble at $\tau = 0$ and then back to the false vacuum at $\tau = \infty$. This path is much like a path through configuration space \vec{x} in multidimensional quantum mechanics parametrized by some path length s . The bubble $\phi_B(r) = \phi_B(\vec{x}, \tau = 0)$ is the analog of the turning point $x = b$.

As the potential is time independent, energy is conserved along the path in time (or Euclidean time). Thus, the potential energy of the bubble is the same as that of the false vacuum $U[\phi_B(r)] = U[\phi_F] = 0$. Since total energy is conserved in τ , we also have

$$U[\phi_B] = \int d^3x \frac{1}{2} (\partial_\tau \phi_B)^2. \quad (5.7)$$

The Euclidean action on the bounce is

$$\begin{aligned} S_E[\phi_B] &= \int d^4x \left[\frac{1}{2} (\partial_\tau \phi_B)^2 + \frac{1}{2} (\vec{\nabla} \phi_B)^2 + V[\phi_B] \right] \\ &= \int_{-\infty}^{\infty} d\tau 2U[\phi_B]. \end{aligned} \quad (5.8)$$

There are a few known exact solutions to Eq. (5.6). An important case is when $V = -\lambda\phi^4$ and the solution has the form

$$\phi_c(\rho) = \sqrt{\frac{2}{\lambda}} \frac{R}{\lambda R^2 + \rho^2} \quad (5.9)$$

for any R and $\lambda > 0$. We call these solutions “quartic bounces.” There are a handful of other known exact solutions [46,49–51]. To explore the relation between the bounce shape and the potential, it is sometimes easier to construct exact potentials given a bounce [by integrating

Eq. (5.6)] than to construct bounces given potentials. Alternatively, one can build up exact solutions perturbatively if there is some small parameter in the potential.

More generally, it is straightforward to solve Eq. (5.6) numerically with the shooting method. The idea is that it is easy to solve differential equations with initial conditions but not with our boundary conditions $\phi'(0) = 0$ and $\phi(\infty) = \phi_F$. So one solves with $\phi'(0) = 0$ and $\phi(0) = \phi_0$ and varies ϕ_0 until the solution satisfies $\phi(\infty) = \phi_F$. If ϕ_0 is too large, the particle will overshoot $\phi = \phi_{FV}$ at large ρ , and if ϕ_0 is too small, it will undershoot. Thus, one iteratively hones in on the right ϕ_0 . One hurdle to rapid convergences with this method is that due to the $\frac{3}{\rho}$ term, one cannot actually start rolling from $\rho = 0$ without a numerical singularity. However, this hurdle is easily surmounted using a trick described in Appendix C.

B. Visualizing $U[\phi]$

As mentioned after Eq. (5.3), the gradient energy present in $U[\phi]$ but not in $V(\phi)$ can mislead our intuition. Unfortunately, because $U[\phi]$ is a functional rather than a function, we cannot simply plot it. More generally, QFT field configuration space, in all its infinite-dimensional glory, can be challenging to visualize. One approach which we now explore is to consider $U[\phi]$ restricted to a one- or two-parameter slice of field space. We can plot $U(\alpha) \equiv U[\phi_\alpha]$ along the parameter(s) α . In such a plot all of the intuition from quantum mechanics applies, including intuition for the tunneling rate coming from the WKB approximation (cf. Sec. IID).

One natural choice for α is the Euclidean time τ parametrizing the path through field space of the bounce. For any bounce $\phi_B(\rho)$, the τ path is given by

$$\phi_\tau(\vec{x}) = \phi_B(\sqrt{r^2 + \tau^2}). \quad (5.10)$$

Now we are using τ as simply a useful coordinate on the slice of field space of interest. For example, with a potential $V(\phi) = -\lambda\phi^4$, the bounce is given by Eq. (5.9), so that

$$U(\tau) = U[\phi_C] = \frac{\pi^2 R^2 \tau^2}{2\lambda(R^2 + \tau^2)^{5/2}}. \quad (5.11)$$

This function is shown for a selection of R in Fig. 14, on the left. We see that while $V(\phi)$ has no barrier, $U[\phi]$ does. Also, from Eq. (5.8) we see that the area under the curves gives the Euclidean action for that bounce. In this case, the areas are all equal as the potential is classically scale invariant and $S_E = \frac{2\pi^2}{3\lambda}$ is independent of R .

Parametrization of paths with τ is fine, but to really think of $U(\tau)$ as a potential, the distance along the path should be properly normalized. Recall that in going from quantum mechanics to quantum field theory, position is replaced by field values $\phi(\vec{x})$. Thus, instead of $ds^2 = dx^2 + dy^2 + \dots$, the measure along a path in field space is given by the somewhat odd-looking line element $ds^2 = \int d^3x [d\phi(x)]^2$. To change from τ to s , we use

$$\left(\frac{ds}{d\tau}\right)^2 = \int d^3x \left(\frac{\partial\phi_B}{\partial\tau}\right)^2 = 2U[\phi_B]. \quad (5.12)$$

Thus, Eq. (5.8) becomes

$$\begin{aligned} S_E &= \int_{-\infty}^{\infty} d\tau 2U[\phi_B] = 2 \int_{-\infty}^0 d\tau 2U[\phi_B] \\ &= 2 \int ds \sqrt{2U[\phi_B]}. \end{aligned} \quad (5.13)$$

The right-hand side of this equation is the exact field-theory analog of the exponent in the WKB formula in Eq. (2.24). The right side of Fig. 14 shows U as a function of s . It is the

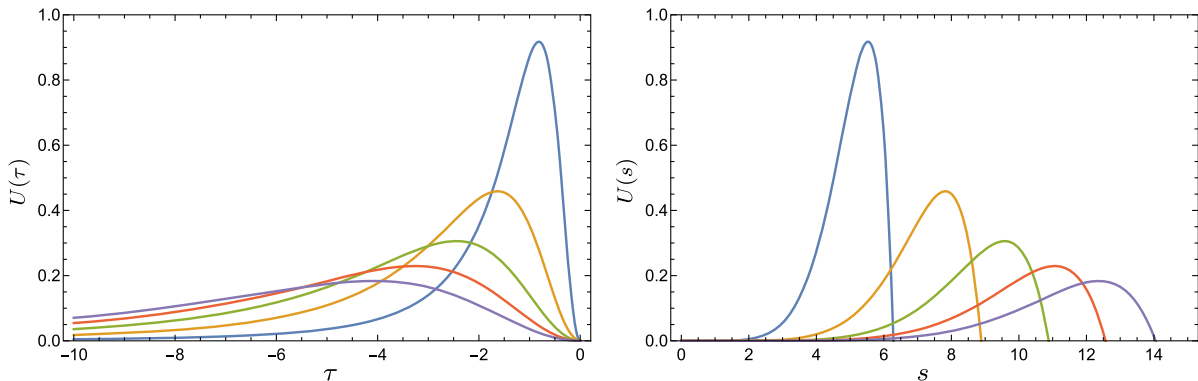


FIG. 14. The proper analog of the potential in QM is the effective energy functional U . This example uses $V(\phi) = -\lambda\phi^4$ with $\lambda = 1$. The left panel shows $U(\tau)$, and the integral of $4U(\tau)$ from $-\infty < \tau < 0$ gives S_E . The right panel shows $U(s)$, and the integral of $\sqrt{2U(s)}$ over $0 < s < \sqrt{\frac{4R}{\lambda}}\pi$ gives S_E . One can see the degeneracy in the fact that each of these barriers has the same integral. Remember that these curves show the height of the barrier in five *different* directions in field configuration space, which happen to all have the same tunneling rate at LO. Curves from steep to shallow have $R = 1, 2, 3, 4$ and 5 .

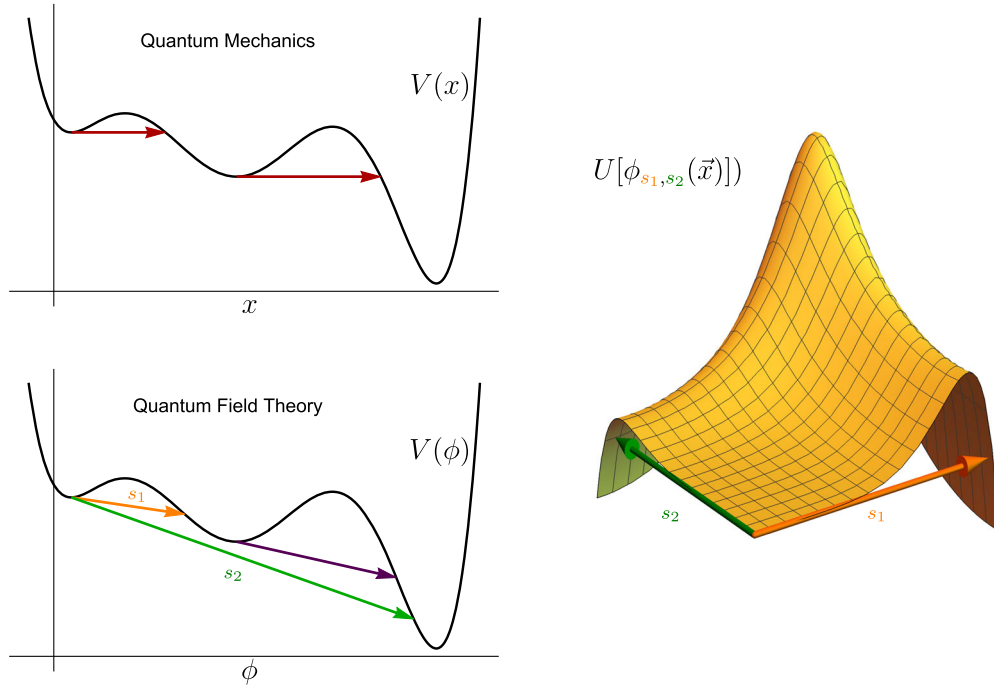


FIG. 15. In quantum mechanics (top left), the tunneling rate is determined by the nearest potential barrier to the closest turning point of $V(x)$. In quantum field theory, tunneling can proceed in different directions of field space. For a potential $V(\phi)$ of the same shape as $V(x)$, tunneling may proceed through the path labeled s_1 corresponding to a family of fields $\phi_{s_1}(\vec{x})$ or through the path labeled s_2 corresponding to a family of fields $\phi_{s_2}(\vec{x})$. The energy functional on a two-parameter family $\phi_{s_1,s_2}(\vec{x})$ is shown on the right.

integral of the square root of these curves which gives the tunneling rate, as in the finite-dimensional WKB formula.

More generally, for any family ϕ_α of field configurations parametrized by some coordinate α , we can find the appropriate normalization using

$$ds^2 = \left[\int d^3x \left(\frac{d\phi_\alpha}{d\alpha} \right)^2 \right] d\alpha^2. \quad (5.14)$$

Once normalized, the decay rate calculated from the minimal action along the subspace of field configurations will approximate the correct decay rate. It will be inexact for two reasons. First, in the fluctuations around the minimum-action path, we will entirely ignore all the modes of fluctuation that are not contained in our subspace. If all we care about is LO accuracy, then this is not a worry. Second, the minimum-action path through our subspace may be very different from the true minimum path, so the dominant e^{-S_E} in the subspace may exponentially overestimate the leading-order decay rate.

Thus, to make our 1D slice *useful*, it should contain the dominant path, as the parametrization by Euclidean time τ , Eq. (5.11), automatically does (since τ is exactly the parameter along the dominant path).

Although we often restrict to 1D slices through field space, it is important to keep in mind that QFT has an infinite variety of field configurations, and potential barriers can be subverted through excursions to large field values.

Indeed, one very important distinction between quantum mechanics and quantum field theory is that in quantum field theory the tunneling rate depends on the potential at arbitrarily large field values. In quantum mechanics, tunneling always proceeds through the closest barrier first, as illustrated in the top-left plot in Fig. 15. In QFT, there can be multiple competing tunneling directions. This is indicated by the orange and green arrows in the lower-left plot of Fig. 15. The end points of the arrows indicate the value of the field at the center of the bubble which forms. Thus, these two arrows correspond to two paths through field space $\phi_{s_1}(\vec{x})$ and $\phi_{s_2}(\vec{x})$, corresponding to two different types of bubbles forming. The barrier (according to $U[\phi]$) along the shorter direction [smaller change in $\phi(0)$] is not necessarily lower than the barrier along the longer direction. In fact, tunneling rates to bubbles with large field values can be exponentially larger than tunneling rates to smaller field values. This is important for the Standard Model, as it makes the stability of our Universe unavoidably sensitive to arbitrarily high-scale physics, as we discuss in Sec. VII C 2.

C. Using approximate solutions

We can visualize the energy along a path through field space as a function of τ if we already know the bounce solution. But what if we do not already have a bounce in hand? Since ϕ_B is an extremum of the action, we can use a

variational approach to get a handle on it by considering a family of approximate solutions.

For most potentials of interest, it is sufficient to consider a two-parameter family of spherically symmetric fields, with the two parameters representing the width of the bubble R which (roughly) controls the gradient energy or surface-tension contribution and the height of the bubble ϕ_0 which we can define as the value at the center $\phi_0 = \phi(r=0)$. For example, a useful two-parameter family includes the “Gaussian” bubbles:

$$\phi_G(\rho) = \phi_0 e^{-\rho^2/R^2}. \quad (5.15)$$

Other simple families are the quartic bubbles, $\phi = \phi_0 \frac{1}{1+\rho^2/R^2}$, and thin-wall bubbles for which $\phi' = 0$ outside of a wall of thickness ℓ .⁶

The Euclidean action for the Gaussian bubbles is

$$S_E[\phi_G] = 2\pi^2 \int d\rho \rho^3 \left[\frac{1}{2} (\partial_\rho \phi_G)^2 + V[\phi_G(\rho)] \right] \quad (5.16)$$

$$= \frac{\pi^2}{2} R^2 \phi_0^2 + a \frac{\pi^2}{4} R^4 V(\phi_0). \quad (5.17)$$

The second term has a potential-dependent number a in it. For smooth potentials we expect $a \sim 1$. For example, if $V = -\lambda\phi^4$, then $a = \frac{1}{4}$.

To find the bounce, we must keep in mind that the bounce is not the true minimum of the action, but only a saddle point. Indeed, the negative-action fluctuation is precisely what is needed to give the path integral the required imaginary part to compute the decay rate. We know a decay can occur if and only if $V(\phi_0) < 0$ for some ϕ_0 , but then S_E in Eq. (5.17) is unbounded from below as $R \rightarrow \infty$. The true minimum is the shot: a path which starts at the false vacuum with enough kinetic energy to make it to the true vacuum quickly, stay there for a long time, and eventually return. Recall that the fluctuation around the bounce with negative eigenvalue (the fluctuation towards the minima) has no nodes and therefore is nonzero at $\tau = 0$. Thus, while the bounce hits the zero-energy surface at $\tau = 0$, the shot or any other spherically symmetric lower-action configuration cannot. So to find the bounce and not the shot, we can minimize the action over ϕ_0 and R , restricting to $U[\phi_G(\tau=0)] = 0$.

⁶This parametrization is more general than the thin-wall approximation originally used by Coleman [40]. The thin-wall approximation requires the two minima to be nearly degenerate so that the bounce remains the true vacuum long enough to neglect the damping term. For potentials which are not nearly degenerate, it has been shown that the thin-wall approximation is not in good quantitative agreement with exact numerical results [52,53]. We will not neglect the damping term. Using Gaussian bubbles seems to give results in good agreement with exact numerical solutions in all the cases we have tried.

The energy of the Gaussian bubbles (at $\tau = 0$) is

$$U = 4\pi \int r^2 dr \left[\frac{1}{2} (\partial_r \phi_G(r))^2 + V(\phi_G(r)) \right] \quad (5.18)$$

$$= \frac{3\pi^{3/2}}{4\sqrt{2}} R \phi_0^2 + b \frac{\pi^{3/2}}{\sqrt{2}} R^3 V[\phi_0]. \quad (5.19)$$

The second term here has another potential-dependent order-one number, b . For example, if $V = -\lambda\phi^4$, $b = \frac{\sqrt{2}}{8}$. Restricting to $U = 0$ means

$$R^2 = -\frac{3}{4b} \frac{\phi_0^2}{V(\phi_0)}. \quad (5.20)$$

Note that $V(\phi_0) < 0$ since there is a friction term that forces the field to start out slightly beyond the turning point where $V = 0$, so $R^2 > 0$. Using this restriction on R , the Euclidean action becomes

$$S_E = \frac{3\pi^2}{64b^2} (8b - 3a) \frac{\phi_0^4}{-V(\phi_0)}. \quad (5.21)$$

More suggestively, if we define the effective quartic as

$$\lambda_{\text{eff}}(\phi) \equiv \frac{V(\phi)}{\phi^4}, \quad (5.22)$$

then the extremum is where

$$\frac{d}{d\phi} \lambda_{\text{eff}} = 0. \quad (5.23)$$

Note that this condition is independent of a and b ; it depends only on the Gaussian bubbles being a reasonable approximation to the true solution.⁷

The above manipulations provide us with a useful shortcut to deduce the approximate features of the bounce simply by looking at the potential: The field value at the center of the bounce ϕ_0 is where λ_{eff} is flat, the size of the bubble is given by Eq. (5.20), and the action on the bounce is no larger than Eq. (5.21). For example, with $V = -\lambda\phi^4$, the exact Euclidean action on the true, quartic bounce is $S_E[\phi_C] = \frac{2\pi^2}{3\lambda} \approx \frac{6.5}{\lambda}$. The extremum along the paths of zero-energy Gaussian bubbles has $S_E[\phi_C] \approx \frac{9.8}{\lambda}$ (although it is something of a degenerate case because the action along the zero-energy Gaussian bubbles is exactly constant).

This effective coupling is commonly used in the literature on the Standard Model effective potential. There it is often invoked in the context of a resummed effective potential where the $\overline{\text{MS}}$ scale μ is set equal to the Higgs

⁷The condition requires $8b > 3a$, which is true for all the examples we have considered.

background field value, and minimizing λ_{eff} is sometimes associated with a renormalization group condition $\partial_\mu \lambda_{\text{eff}}(\mu) = 0$. Here we see that the relevance of λ_{eff} to the bounce solutions has nothing to do with renormalization: It follows generically in classical potentials.

Another fun exercise with these Gaussian bubbles is to determine how R would depend on time in a classical field theory. Consider a Gaussian bubble

$$\phi_G(r, t) = \phi_0 \exp \left[-\frac{r^2}{R(t)^2} \right]. \quad (5.24)$$

To determine the time dependence, we can integrate over d^3x in the Minkowski-space action to get an effective one-dimensional Lagrangian:

$$L = 4\pi \int dr r^2 \left[\frac{1}{2} (\partial_t \phi_G)^2 - \frac{1}{2} (\partial_r \phi_G)^2 - V(\phi_G) \right] \quad (5.25)$$

$$= \frac{15\pi^{3/2}}{16\sqrt{2}} R \dot{R}^2 \phi_0^2 - \frac{3\pi^{3/2}}{4\sqrt{2}} R \phi_0^2 - b \frac{\pi^{3/2}}{\sqrt{2}} R^3 V(\phi_0), \quad (5.26)$$

where b is the same order-one constant as in Eq. (5.19). The last two terms in this expression for L are just $-U$, from Eq. (5.19). Calculating the Euler-Lagrange equations from this Lagrangian gives

$$\ddot{R} = -\frac{\dot{R}^2}{2R} - \frac{8bV(\phi_0)R}{5\phi_0^2} - \frac{2}{5R}. \quad (5.27)$$

We immediately confirm that for bubbles to grow $V(\phi_0)$ must be negative (as expected). For bubbles at rest ($\dot{R} = 0$) the condition for bubble growth ($\ddot{R} > 0$) is that

$$R^2 > -\frac{\phi_0^2}{4bV(\phi_0)}. \quad (5.28)$$

Comparing to Eq. (5.20), we see that these bubbles grow if $R^2 > \frac{1}{3} R_{E=0}^2$, where $R_{E=0}$ is the condition for the bubble to have zero energy. In particular, Gaussian bubbles which form on the $U = 0$ surface Σ do indeed grow with time.

VI. TUNNELING AT NLO AND EFFECTIVE ACTIONS

The previous sections have discussed tunneling in quantum mechanics and quantum field theory, both through an exact, nonperturbative definition of the decay rate and through the perturbative approximation to the rate coming from expanding around saddle points. All of these calculations assumed that a potential $V(\phi)$ of a single scalar field ϕ was given. In quantum field theory, the physics of tunneling is often associated with an effective potential V_{eff} , generated by integrating out quantum corrections. In some cases, the instability is even induced from radiative

corrections. One might imagine that the effective potential can simply be used in place of the classical potential in the tunneling formulas. However, this is incorrect; it will overcount the quantum corrections.

A classic example where the instability is a quantum effect is the Coleman-Weinberg potential [10]. Coleman and Weinberg considered the theory of a complex scalar field and an Abelian gauge boson, i.e., massless scalar QED. In the Coleman-Weinberg model, the classical potential $V(\phi) = \lambda|\phi|^4$ is scale invariant and has an absolutely stable minimum at $\phi = 0$ (for $\lambda > 0$). In contrast, the effective potential has a minimum at $|\phi| > 0$, and thus $\phi = 0$ is unstable. The scale for the new minimum is determined by dimensional transmutation from the renormalization group scale of the couplings. What is the correct procedure to compute the tunneling rate in this model? One cannot use $V(\phi)$ in the tunneling rate formulas above since there are no bounce solutions for this potential to expand around. Nor can one use V_{eff} , as the quantum fluctuations have already been integrated out, so there is no longer a path integral.

The same problem occurs in the Standard Model, where the classical potential for the Higgs field has an absolute minimum at the electroweak vacuum expectation value $\langle h \rangle = 246$ GeV. The effective potential indicates tunneling to a very high scale. However, how can we calculate this tunneling rate accurately?

A. Effective actions

First, let us quickly review what is meant by the terms effective potential and effective action. Unfortunately, these same terms are used for a few different but related functions.

The simplest way to compute an effective action for a field ϕ is to integrate out (perform the path integral over) all the *other* fields in the theory. For example, with two fields ϕ and χ , one could integrate out χ to get an effective action for ϕ :

$$e^{-S_{\text{eff}}[\phi]} \equiv \int \mathcal{D}\chi e^{-S[\phi, \chi]}. \quad (6.1)$$

The path integral on the right is to be calculated for a fixed (but possibly position-dependent) background configuration ϕ . In particular, no loops involving virtual ϕ particles are to be included in the calculation on the right-hand side. Although computing effective actions this way for general field configurations ϕ in any theory is essentially impossible, a momentum expansion is feasible. For example, the Euler-Heisenberg action is an effective action of this type, where the electron is integrated out in QED and the background A^μ is assumed constant.

Even if the classical action for ϕ , $S[\phi, 0]$, has no instability but the effective action computed this way does, one can proceed to calculate the tunneling rate using the

effective action. There is no double counting because the fluctuations of some fields [χ in Eq. (6.1)] are included in the calculation of the effective action while the fluctuations of other fields (ϕ) are only included when the rate is computed, as in Eq. (5.4). This approach was explored by Weinberg in [54]. In particular, Weinberg observed that in scalar QED integrating out A^μ generates the instability. He also observed that one can additionally integrate out the imaginary part of ϕ , leaving an effective action that depends only on the real part of ϕ . A derivative expansion, justified with $\lambda \sim e^4$ scaling, was critical to Weinberg's argument, as we review below.

It may turn out that the fluctuations of the field ϕ itself are required to generate the instability. In that case, one cannot simply integrate out ϕ , or there is nothing left for the effective action to depend on. For example, one might try to replace $\phi \rightarrow \phi + \bar{\phi}$, with $\bar{\phi}$ a fixed external background field, and then integrate out ϕ . That is, we can compute

$$Z[0] = \int \mathcal{D}\phi e^{-S[\phi]} = \int \mathcal{D}\phi e^{-S[\phi + \bar{\phi}]}. \quad (6.2)$$

Since the path integral is invariant under field reparametrizations, $Z[0]$ computed this way is just a number. That is, it will not depend on $\bar{\phi}$ at all. Nevertheless, this is the same Z used in the potential-deformation method for computing the tunneling rate. Thus, one should expect $\text{Im} \ln Z[0]$ to be related to the decay rate. Note that even if the classical potential has no instability, as long as the full theory has an instability, one can still, in principle, compute the decay rate in this way. However, one cannot use the saddle-point approximation since we have no classical bounce solution.

The usual procedure for computing an effective action that admits a bounce is to begin with the generating functional

$$W[J] = -\ln Z[J] = -\ln \int \mathcal{D}\phi e^{-S[\phi] - \int d^4x \phi(x) J(x)} \quad (6.3)$$

and take its Legendre transform

$$S_{\text{eff}}[\bar{\phi}] = W[J_{\bar{\phi}}] - \int J_{\bar{\phi}}(x) \bar{\phi}(x). \quad (6.4)$$

Here, $J_{\bar{\phi}}$ is defined so that $\frac{\delta W}{\delta J}|_{J=J_{\bar{\phi}}} = \bar{\phi}$. Since $\frac{\delta W}{\delta J}|_{J=J_{\bar{\phi}}}$ is the expectation value of the field in the quantum theory, this says that $J_{\bar{\phi}}$ is the background current required to make the expectation value of $\phi = \bar{\phi}$. Since $\frac{\delta S_{\text{eff}}}{\delta \bar{\phi}} = -J_{\bar{\phi}}$, we conclude that the true vacuum of the theory $\bar{\phi}_{0,\text{TV}}$ is given by an extremum of S_{eff} .

Formally, the Legendre transform requires that there is a one-to-one correspondence between $J_{\bar{\phi}}$ and $\bar{\phi}$. That is, the functionals $W[J]$ and $S_{\text{eff}}[\bar{\phi}]$ must be convex for the

Legendre transform to exist. However, convex actions never admit tunneling.

The incompatibility of convexity and tunneling is not as problematic as it sounds. First of all, in practice, one computes the effective action not through the Legendre transform but through the computation of 1PI diagrams with background fields. The action generated by background fields does not have to be convex (in the Standard Model, for example, it is not). Weinberg and Wu have argued [55] that this 1PI effective action is more physical than the Legendre transform version since it does not include linear superpositions of vacua. Although their argument is compelling, it does not help if we want to use the Legendre transform definition to compute the tunneling rate.

A more convincing argument is that in the potential-deformation method, we actually *do* want a convex potential, whose single minimum corresponds to the false vacuum. We can get this by deforming the potential or, more directly, by sticking to a steepest-descent contour. Following the logic from Sec. III, we can define the effective action $S_{\text{eff}}[\bar{\phi}_0]$ through the Legendre transform along the convex steepest-descent contour through the false vacuum. We then write the imaginary part as the discontinuity of the effective action defined along the steepest-descent contour through the bounce. So

$$\frac{1}{2} \frac{\Gamma}{V} = \frac{1}{2} \text{Im} \left(\frac{1}{\mathcal{T}V} e^{-S_{\text{eff}}[\bar{\phi}_0]} \right) \quad (6.5)$$

where $\bar{\phi}_0$ is the bounce solution to the equation of motion, $S'_{\text{eff}}[\bar{\phi}_0] = 0$. In principle, we should normalize this by $e^{-S_{\text{eff}}[0]}$, but this is equal to 1 by our choice of origin $\phi = 0$. The requirement to only include the bounce saddle point is built into the way we compute in perturbation theory; we only include small fluctuations around the argument of S_{eff} . The gauge dependence of this way of computing the tunneling rate, particularly with regard to boundary conditions, is discussed in [56].

Let us briefly review how one evaluates the one-bounce contribution to the path integral Z (see Sec. VI B) using the functional determinant and the classical action. There $Z = K e^{-S[\phi_c]}$ at NLO, where $K \equiv \left(\frac{\text{Det}[S''[\phi_c]]}{\text{Det}[S''[0]]} \right)^{-\frac{1}{2}}$. Due to the zero modes from translations, $K = \mathcal{T} V K'$ where K' has the determinants evaluated with the zero modes removed, and $\mathcal{T} V$ is the volume of Euclidean space-time. Note that K' has units of mass,⁴ and it must come from some characteristic length scale R from the bounce solution; thus, we can reexpress this as $K' \sim \frac{1}{R^4}$ up to some dimensionless number we expect to be of order one [19].

We expect the same thing to happen for the effective action using Eq. (6.5), but to our knowledge, no complete proof exists. The bounce solution $\bar{\phi}_0$ is space-time translation invariant, so Z should be proportional to $\mathcal{T} V$. To get

the right dimensions, we must compensate with some characteristic scale R from the bounce solution. We then find

$$\frac{\Gamma}{V} = \frac{1}{R^4} \text{Im}[e^{-S_{\text{eff}}[\bar{\phi}_0]}]. \quad (6.6)$$

We will show that the tree-level and logarithmic terms agree between the effective potential method and the functional determinant method in the next section. We will also see that if we wish to check the prefactor of Eq. (6.6) and the $\mathcal{O}(1)$ terms in the exponent, we would have to know all the higher derivative terms in the effective action.

In summary, in situations where the classical potential is stable but there is an instability in the quantum theory, the tunneling rate is given by the exponential of the effective action evaluated on the solution $\bar{\phi}_0$ to its equations of motion. Of course, the effective action can be used to calculate the decay rate even in situations where the classical potential admits tunneling. However, this approach is only useful to the extent that the effective action can be computed exactly, or in some approximation consistent with a perturbative expansion of the decay rate. Next, we explore what the effective action looks like and whether it is useful for computing tunneling rates in an example.

B. NLO tunneling in scalar field theory

Consider the theory of a real scalar field with classical potential $V(\phi) = \lambda\phi^4$. For $\lambda < 0$, the vacuum at $\phi = 0$ is unstable. The Euclidean equations of motion are solved by the quartic bounces in Eq. (5.9): $\phi_C(\rho) = \sqrt{\frac{2}{|\lambda|}} \frac{R}{R^2 + \rho^2}$. The tree-level Euclidean action evaluated on the quartic bounce is $S_E[\phi_C] = \frac{2\pi^2}{3|\lambda|}$, independent of the bubble size R . The decay rate at NLO can be computed either using Eq. (6.2), summing over the Gaussian fluctuations around ϕ_C , or using Eq. (6.6) where the field fluctuations around an arbitrary background field configuration contribute to the form of the effective action.

To compute the rate by integrating over Gaussian fluctuations around ϕ_C , we follow the approach of [15,57]. We directly calculate

$$\frac{\Gamma}{V} = \frac{S_E[\phi_C]^2}{4\pi^2} \left[\frac{-\det' S_E''[\phi_C]}{\det S_E''[0]} \right]^{-1/2} e^{-S_E[\phi_C]}. \quad (6.7)$$

Here \det' refers to the functional determinant where the zero modes corresponding to translations have been removed. This functional determinant can be computed by solving the eigenvalue equation $[-\partial^2 + W(\rho)]\phi = \lambda\phi$ and taking the product of the eigenvalues. We relegate details of this calculation to Appendix E. In $\overline{\text{MS}}$, the result is that

$$\frac{\Gamma}{V} = \frac{1}{R^4} \exp \left[-\frac{2\pi^2}{3|\lambda(\mu)|} + 3 \ln(R\mu) + \mathcal{O}(1) \right] \quad (6.8)$$

where $\mathcal{O}(1)$ is some order-one number coming from the evaluation of the functional determinant, and R takes a specific value which saturates the path integral over the associated collective coordinate.

As a quick check, we can verify that the rate is independent of μ to order λ , using $\beta_\lambda = \mu \frac{\partial}{\partial \mu} \lambda = \frac{9}{2\pi^2} \lambda^2$. Note that the NLO rate breaks the degeneracy in R . In fact, for a given value of μ , the rate appears to go to zero if $R \rightarrow 0$, or the rate is unbounded from above if $R \rightarrow \infty$. As the rate increases, however, the logarithm also grows to the point where subleading orders become relevant. The logarithms are minimized at the scale μ where $\beta_\lambda = 0$. This is consistent with the prescription for finding the bubble shape for a general potential discussed in Sec. VC.

The second method, using Eq. (6.6), requires knowing the exact effective action, so that its equations of motion can be solved and its solution used to evaluate the rate. If we write $\bar{\phi}_0 = \phi_C + \hbar\phi_1 + \dots$ and $S_{\text{eff}} = S_0 + \hbar S_1 + \dots$, then

$$S_{\text{eff}}[\bar{\phi}_0] = S_0[\phi_C] + \hbar S'_0[\phi_C] + \hbar S_1[\phi_C] + \mathcal{O}(\hbar^2). \quad (6.9)$$

Since $S'_0[\phi_C] = 0$, to NLO we only need to compute $S_1[\phi_C]$. That is, the corrections to the shape of the bubble come in first at next-to-next-to-leading order. In particular, we do not have to compute or solve the equations of motion for the effective action (which could have been very difficult considering that S_{eff} is nonlocal).

Computing the full effective action $S_{\text{eff}}[\phi]$ is essentially impossible, even at one loop. We can, however, compute it order by order in a momentum expansion. The leading contribution, with no derivatives, is the effective potential:

$$V_{\text{eff}} = \lambda\phi^4 + \frac{9}{4\pi^2} \lambda^2 \phi^4 \left(\ln \frac{12\lambda\phi^2}{\mu^2} - \frac{3}{2} \right). \quad (6.10)$$

This is computed in the background field method assuming the background fields are constant. To get the terms with two derivatives, we compute diagrams with background fields with nonzero momenta and take two derivatives with respect to those momenta, then set the momenta to zero. More details are given in Appendix D. The result is that to one loop with up to four derivatives, the effective action in Euclidean space is

$$\begin{aligned}
S_{\text{eff}}[\phi] = \int d^4x \left[\lambda \phi^4 + \frac{9}{4\pi^2} \lambda^2 \phi^4 \left(\ln \frac{12\lambda \phi^2}{\mu^2} - \frac{3}{2} \right) \right. \\
+ \frac{1}{2} (\partial_\mu \phi)^2 \left(1 + \frac{1}{4\pi^2} \lambda \right) - \frac{1}{2} (\Box \phi)^2 \frac{1}{480\pi^2} \frac{1}{\phi^2} \\
+ \frac{1}{2} (\partial_\mu \phi)^2 \Box \phi \frac{1}{720\pi^2} \frac{1}{\phi^3} \\
\left. - \frac{1}{8} (\partial_\mu \phi)^2 (\partial_\nu \phi)^2 \frac{1}{360\pi^2 \phi^4} + \mathcal{O}(\partial^6) \right]. \quad (6.11)
\end{aligned}$$

We can now check whether the momentum expansion is justified for use in the calculation of the decay rate.

The effective potential (zero-derivative terms) contributes to the action as

$$\begin{aligned}
S_{\text{eff}}^V[\phi_C] &= 2\pi^2 \int d\rho \rho^3 V_{\text{eff}}[\phi_C(\rho)] \\
&= -\frac{2\pi^2}{3|\lambda|} - \frac{19}{4} - 3 \ln \frac{R\mu}{2\sqrt{6}} \pm i \frac{3\pi}{2} \quad (6.12)
\end{aligned}$$

where $\pm i \frac{3\pi}{2}$ comes from $\ln \frac{\lambda}{|\lambda|}$ and $\lambda < 0$. This imaginary term in the action makes $e^{-S_{\text{eff}}}$ imaginary since $e^{\pm i \frac{3\pi}{2}} = \mp i$. The sign is ambiguous, but we pick the sign such that $\Gamma > 0$. The two-derivative terms

$$S_{\text{eff}}^{2\text{der}}[\phi] = \int d^4x \left(1 + \frac{\lambda}{4\pi^2} \right) \frac{1}{2} (\partial_\mu \phi)^2 \quad (6.13)$$

contribute to the action on the bounce as

$$S_{\text{eff}}^{2\text{der}}[\phi_C] = \frac{4\pi^2}{3|\lambda|} - \frac{1}{3}. \quad (6.14)$$

Using Eq. (6.6), and ignoring the $\mathcal{O}(1)$ numbers, we find

$$\frac{\Gamma}{V} = \frac{1}{R^4} \text{Im} \left[e^{-i \frac{3\pi}{2}} e^{-\frac{2\pi^2}{3|\lambda|} + 3 \ln(R\mu) + \mathcal{O}(1)} \right] \quad (6.15)$$

$$= \frac{1}{R^4} \exp \left[-\frac{2\pi^2}{3|\lambda|} + 3 \ln(R\mu) + \mathcal{O}(1) \right]. \quad (6.16)$$

We see that the logarithm and $\frac{1}{\lambda}$ term in Eq. (6.8) are reproduced exactly. This of course has to happen since all the μ dependence at one loop must be compensated by the potential and kinetic terms in the classical action.

Note that the NLO part of Eq. (6.14) (the $\frac{1}{3}$) has the same scaling as the NLO contributions from the potential, Eq. (6.12). Thus, adding derivatives does not seem to give additional suppression.

To understand whether higher order terms give additional suppression, let us turn to the four-derivative terms in the one-loop effective action (see Appendix D):

$$\begin{aligned}
S_{\text{eff}}^{4\text{der}} &= -\frac{1}{2} (\Box \phi)^2 \frac{1}{480\pi^2} \frac{1}{\phi^2} + \frac{1}{2} (\partial_\mu \phi)^2 \Box \phi \frac{1}{720\pi^2} \frac{1}{\phi^3} \\
&\quad - \frac{1}{8} (\partial_\mu \phi)^2 (\partial_\nu \phi)^2 \frac{1}{360\pi^2} \frac{1}{\phi^4}. \quad (6.17)
\end{aligned}$$

Since the bounce $\phi_C \sim \frac{1}{\sqrt{|\lambda|}}$, when we evaluate the action on the bounce, all these terms will scale like λ^0 . Thus, they will be of the same order as the potential and two-derivative terms.

More generally, one-loop terms with any number of derivatives contribute to $S_{\text{eff}}[\phi_C]$ at the same order as the one-loop effective potential. To see this, note that whenever a factor of momentum is pulled out of a loop graph, it must be compensated by an effective mass, by dimensional analysis. In $\lambda \phi^4$ theory, the effective mass is $m_{\text{eff}}^2 = 12\lambda \phi^2$. Since $\phi_C \sim \frac{1}{\sqrt{\lambda}}$, adding two derivatives and $\frac{1}{m_{\text{eff}}^2}$ does not change the power counting.

We conclude that the derivative expansion is *not* justified for calculating decay rates. That is, to compute the decay rate at NLO in $\lambda \phi^4$ theory using the effective action, we need the complete effective action to one loop, not just the leading-momentum-dependent terms. Since this effective action is nearly impossible to compute, this method is not feasible for computing NLO decay rates in quantum field theory. Similar conclusions were reached through a calculation in quantum mechanics in [58].

Keep in mind that we have only shown that the derivative expansion does not work for $\lambda \phi^4$ theory. In other theories, it may be useful, but it depends on the circumstances. For example, in the Coleman-Weinberg model, there are two couplings, λ and e . For e small, the running of the couplings is perturbative and λ goes from $-\infty$ to $+\infty$ while e remains small. Thus, the theory is completely specified by a single small number e . The loop corrections in this theory induce spontaneous symmetry breaking at a scale μ where $\lambda(\mu) \sim e(\mu)^4$. Having $\lambda \sim e^4$ is more than just a numerical association: One must power count with this scaling so that physical quantities, such as the vector-to-scalar mass ratio, are gauge invariant [59]. In the Coleman-Weinberg model, the effective masses relevant at one loop are of the form $m_{\text{eff}} \sim e\phi$. When $\lambda \sim e^4$, then $\frac{\Box}{m_{\text{eff}}^2} \sim \frac{\lambda}{e^2} \sim e^2$, so higher derivative terms are power suppressed.

Weinberg and Metaxas [60] used the $\lambda \sim e^4$ power counting to show that the tunneling rate in the Coleman-Weinberg model is gauge invariant at NLO. But although $\lambda \sim e^4$ scaling is critical for spontaneous symmetry breaking, it is not the appropriate scaling for the calculation of tunneling rates. As discussed earlier in this section, the scale appropriate for tunneling is where $\beta_\lambda(\mu) = 0$. Since $\beta_\lambda \sim e^4 + e^2 \lambda + \lambda^4$, $\beta_\lambda = 0$ is compatible with the normal loop power counting where $\lambda \sim e^2$. With this counting,

$m_{\text{eff}} \sim 1$ as in $\lambda\phi^4$ theory and $\frac{\square}{m_{\text{eff}}}$ corrections are as important as corrections to the effective potential.

VII. STANDARD MODEL

In the previous sections, we have discussed general features of tunneling calculations in quantum mechanics and quantum field theory. In this section, we apply some of those insights to the Standard Model. In particular, we consider the question of how to calculate the lifetime of the metastable Standard Model vacuum with $\langle h \rangle = 246$ GeV in a systematically improvable way. We also discuss the UV sensitivity of the decay rate calculation.

A. Effective potential and gauge invariance

A discussion of vacuum stability in the Standard Model usually begins with the Standard Model effective potential. As we have seen, this is not all we need to accurately describe tunneling. This potential $V_{\text{SM}}(h, \mu)$ is a function of a constant real scalar background Higgs field h , the renormalization group scale μ , and the various couplings in the theory. It has been computed to two-loop accuracy with resummation of large logarithms to the three-loop level.

An important feature of the Standard Model effective potential is that it is nearly scale invariant. The only mass scale in the Standard Model is the single dimensionful parameter v (the Higgs vev). For $h \gg v$, we can set $v = 0$, along with the Higgs mass and all other masses. Then all of the scale dependence comes from quantum corrections and dimensional transmutation. To make this clear, we often write

$$V_{\text{SM}}(h, \mu) = \frac{1}{4} h^4 \lambda_{\text{eff}}(h, \mu). \quad (7.1)$$

The function $\lambda_{\text{eff}}(h, \mu)$ is dimensionless, matches the Higgs quartic at the weak scale, and changes slowly (logarithmically) as h is increased. Here μ is the $\overline{\text{MS}}$ scale, and all of the μ dependence is either implicit, through the gauge couplings $g_i(\mu)$, top Yukawa $\lambda_t(\mu)$, etc., or explicit in terms of the form $\ln \frac{h}{\mu}$. The explicit and implicit μ dependences are related by the renormalization group equation

$$\left(\mu \frac{\partial}{\partial \mu} - \gamma h \frac{\partial}{\partial h} + \beta_i \frac{\partial}{\partial \lambda_i} \right) V_{\text{SM}}(h, \mu) = 0 \quad (7.2)$$

with γ the Higgs field anomalous dimension.

It is commonplace to resum the effective potential by setting $\mu = h$ [61]. Indeed, this is the natural choice, as h is the only scale around. However, setting $\mu = h$ is dangerous. For example, if one is interested in extrema of V_{SM} , then solving $\frac{\partial}{\partial h} V_{\text{SM}}(h, \mu) = 0$ and setting $\mu = h$ afterwards does not give the same value of h as solving $\frac{\partial}{\partial h} V_{\text{SM}}(h, h) = 0$. Or, to calculate the tunneling rate, one must evaluate the effective action on a solution to the equations of motion before setting $\mu = h$, rather than after.

Another issue to keep in mind is that V_{SM} is not gauge invariant. In fact, explicit gauge dependence is *required* for Eq. (7.2) to hold since the anomalous dimension γ is gauge dependent. Moreover, not all the gauge dependence can be associated with γ ; there is further gauge dependence in the nonlogarithmic terms in V_{SM} as well. The gauge dependence of effective potentials is fairly well understood. Recall that the effective potential describes the energy of the system in the presence of a background current J . Since this current is a charged source, it depends on gauge, and therefore the potential will be gauge dependent whenever $J \neq 0$. This explanation also implies that at the extrema, where $J = 0$, the effective potential is gauge invariant. Thus, the energy of any metastable minimum is gauge invariant and possibly physical. This is indeed true and has been checked explicitly. A more rigorous proof that the vacuum energy is gauge invariant relies on the Nielsen identity [62] (see also [63]).

Since the tunneling rate is physical, it should be gauge invariant. In fact, as it appears to be given by the Euclidean effective action at its extremum (the bounce), its gauge invariance also follows from the Nielsen identity. Some subtleties in establishing gauge invariance, involving boundary conditions on the path integral, were recently explored in [56] (see also [18]). Unfortunately, we know of no explicit demonstrations of gauge invariance, say, at next-to-leading order, even in scalar QED.⁸ Such demonstrations may have practical implications for tunneling rate calculations as they had for absolute stability bounds.

In [64], it was shown that for the absolute stability bound in the Standard Model to be gauge invariant, two modifications of the usual procedure were necessary. To be clear, by “usual procedure” we mean using the resummed effective potential with $\mu = h$ in Landau gauge as in [61]. The modifications were as follows:

1. The effective potential had to be known to fixed order in \hbar assuming that $\lambda \sim \hbar$. This is the same scaling as introduced by Coleman and Weinberg, but it had not been used in the SM.
2. The renormalization group equation for the effective potential should not be solved, and we should not set $\mu = h$; rather, the couplings run to a scale μ_X where the leading-order (in the modified \hbar scaling) potential is minimized and the potential evaluated at fixed order.

Let us now consider whether these two modifications also must be applied to the calculation of tunneling rates.

First, we would like to know if we should use $\lambda \sim \hbar$ as in the absolute stability calculation, or $\lambda \sim \hbar^0$ as in ordinary perturbation theory. The reason the modified scaling is appropriate for absolute stability is because an extremum of

⁸In [60], Weinberg and Metaxas confirm the Nielsen identity at one loop using $\lambda \sim \hbar$ scaling, but they do not explicitly show gauge invariance of the one-loop rate.

the potential requires that the growth of the monotonic classical potential $V = \frac{1}{4}\lambda h^4$ be canceled by loop effects scaling like \hbar . This is only possible if λ is anomalously small, $\lambda \sim \hbar$. For absolute stability, we need the potential in only a small neighborhood of an extremum, so it is consistent for $\lambda \sim \hbar$ within the entire neighborhood. However, for tunneling, which involves the potential connecting different extrema, we necessarily need a large range of scales, and it is inconsistent for $\lambda \sim \hbar$ throughout that range. It is easy to see that in the Standard Model, $\lambda(\mu)$ is indeed not anomalously small away from extrema of V_{SM} . We conclude that for a tunneling rate, one should not take $\lambda \sim \hbar$, but instead take $\lambda \sim \hbar^0$ and use ordinary perturbation theory. As discussed in Sec. VIB, this means that all higher derivative terms in the Standard Model effective action will be as important as corrections to the effective potential.

For the second point, we have a new problem: The effective potential describes a minimum near $h = 0$ (our vacuum), a maximum where the Landau-gauge field is around $h \sim 10^9$ GeV, and another minimum at around $h \sim 10^{30}$ GeV. These scales are so far apart that there are necessarily large logarithms in the effective potential and presumably resummation is critical. Indeed, with $\mu \approx 200$ GeV, the resummed effective potential is not at all well described by fixed-order perturbation theory. But if we require resummation, which mixes orders in \hbar , checking gauge invariance order by order in \hbar is impossible.

To investigate further, we write the SM effective potential to two-loop order (in Landau gauge) as

$$V_{\text{SM}}^{(2)}(h) = h^4 \left(A + B \ln \frac{h}{\mu} + C \ln^2 \frac{h}{\mu} \right). \quad (7.3)$$

Here, A , B , and C are calculable functions of the SM couplings, which in turn depend on μ . A lot of this μ dependence is canceled by the explicit μ dependence in Eq. (7.3), but not all of it. At tree level with $\mu = 200$ GeV, $A = \frac{\lambda}{4} = 0.031$ and $B = C = 0$. Values for A , B , and C at two loops for different choices of μ are given in Table I.

Conveniently, we find that there is good agreement between the traditional resummed potential $V_{\text{SM}}(h, h)$ and the fixed-order potential $V_{\text{SM}}^{(2)}$ when $\mu = 10^{17}$ GeV as can be seen in Fig. 16. By good agreement, we mean that the extrema are fairly close. Both have maxima at around $h = 10^9$ GeV and minima around $h = 10^{30}$ GeV. (In contrast, taking $\mu = 10^{10}$ GeV leads to a maximum at $h = 10^{10}$ GeV and a minimum at 10^{18} GeV.)

In conclusion, we have shown that fixed-order perturbation theory can be used to calculate the Standard Model effective potential in good quantitative agreement with the resummed potential. This implies that there is no impediment in trying to establish *explicitly* the gauge independence of the tunneling rate in the Standard Model using the

TABLE I. Values of A , B , and C in the fixed-order SM effective potential $V_{\text{SM}}^{(2)}(h) = h^4(A + B \ln \frac{h}{\mu} + C \ln^2 \frac{h}{\mu})$ as in Eq. (7.3) for various choices of μ in Landau gauge. We have used $m_h^{\text{pole}} = (125.14 \pm 0.24)$ GeV and $m_t^{\text{pole}} = (173.34 \pm 1.12)$ GeV. For $\mu = 10^{17}$ GeV, the two-loop potential is in good agreement with the resummed potential.

μ (GeV)	A	B	C
200	0.0387	-0.00777	8.38×10^{-4}
10^5	0.151	-0.00248	1.64×10^{-4}
10^{10}	0.000751	-6.08×10^{-4}	2.84×10^{-5}
10^{15}	-0.00313	-1.48×10^{-4}	9.32×10^{-6}
10^{17}	-0.00352	-6.19×10^{-5}	6.76×10^{-6}
10^{20}	-0.00347	2.28×10^{-5}	4.86×10^{-6}

fixed-order effective action if one can calculate all the higher derivative corrections relevant to that order. Even in the absence of such an explicit demonstration, it seems reasonable to expect that, since resummation mixes up orders in perturbation theory, one should always use the fixed-order potential and never the resummed one.

B. SM tunneling rate

The Standard Model effective potential has a minimum at our electroweak vacuum ($h \sim 0$) with $V_{\text{SM}}(0) = 0$ and a maximum value of $V_{\text{max}} \approx (10^{10} \text{ GeV})^4$. It then turns over and runs negative, eventually hitting another minimum at around $V_{\text{min}} \sim -(10^{30} \text{ GeV})^4$ [64]. The field values where these extrema are taken are gauge dependent, but the energy densities at the extrema are gauge independent. One should not take the value 10^{30} GeV for the scale of the absolute minimum very seriously: Quantum gravity will obviously modify this scale, as we discuss below in Sec. VII C.

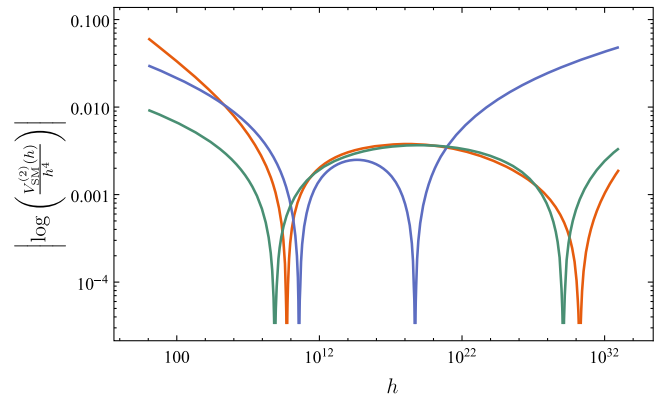


FIG. 16. The effective potential in the Standard Model in Landau gauge resummed with $\mu = h$ (red curve) is compared to the effective potential at two loops with $\mu = 10^{10}$ GeV (blue) and $\mu = 10^{17}$ GeV (green).

The fact that the potential grows, has a maximum, and then decreases to a minimum indicates that there is a potential barrier through which the Higgs field can tunnel. However, as we saw in Sec. V, this is not the correct picture. It is not the barrier of the effective potential of height 10^{10} GeV that we should think about tunneling through. Instead we must study the barrier in the potential energy functional $U[h]$ defined in Eq. (5.3). To see this, recall the general result from Sec. VC that for a potential $V(h)$ the bubble size is determined by the condition $\frac{d}{dh} \frac{V(h)}{h^4} = 0$. For the Standard Model, this scale is $h \sim 10^{17}$ GeV. As we can see from Table I, near the scale $\mu = 10^{17}$ GeV, the Standard Model effective potential is basically just

$$V_{\text{SM}}(h \sim 10^{17} \text{ GeV}) \approx \frac{\lambda_0}{4} h^4, \quad \frac{\lambda_0}{4} = -0.00352. \quad (7.4)$$

Here λ_0 is approximately the value of the Higgs quartic at its minimum, $\lambda_0 \approx \lambda(\mu = 10^{17})$.

There are solutions to the Euclidean equations of motion following from Eq. (7.4) for any bubble size R . These are just the quartic bounces as in Eq. (5.9): $h_C^R(\rho) = \sqrt{\frac{8}{|\lambda_0|}} \frac{R}{R^2 + \rho^2}$, and the energy functional U for the different bubble shapes is shown in Fig. 14. The tree-level action on any of these bounces is $S[h_C^R] = \frac{8\pi^2}{3|\lambda_0|}$. The decay rate per unit volume is therefore

$$\frac{\Gamma^{\text{LO}}}{V} = \exp\left(-\frac{8\pi^2}{3|\lambda_0|}\right) \approx 10^{-812} \quad (7.5)$$

independent of R . The right-hand side has the wrong units, as the units come in at NLO. However, the rate is so small that to a first approximation, the units do not even matter. The difference between choosing a scale of the Planck mass and a scale of the size of the Universe (to the fourth power) gives only a factor around 10^{200} .

If the degeneracy in R persists to all orders, it would invalidate the method of computing the tunneling rate in the Gaussian approximation: Varying R would lead to a zero eigenvalue of the quadratic fluctuations, and it would be unsuppressed in the path integral. We could possibly treat the scale invariance as we treat translation symmetries: integrate over R as a collective coordinate to produce a “volume” factor. However, the R dependence is in fact broken by loop corrections.

There is a shortcut to determining the R dependence of the NLO decay rate. The decay rate, being physical, satisfies a renormalization group equation of the form

$$\left(\mu \frac{\partial}{\partial \mu} + \beta_i \frac{\partial}{\partial \lambda_i}\right) \Gamma(R, \mu) = 0. \quad (7.6)$$

The μ dependence is known from the SM β functions, and the R dependence is determined by dimensional analysis. Explicitly, we must have

$$\begin{aligned} \frac{1}{V} \Gamma^{\text{NLO}} = R^{-4} \exp \left[-\frac{8\pi^2}{3|\lambda_0|} - \left(-4 + 2 \frac{\lambda_t^2}{|\lambda_0|} - \frac{2g_2^2 + g_Z^2}{2|\lambda_0|} \right. \right. \\ \left. \left. + \frac{\lambda_t^4}{\lambda_0^2} - \frac{2g_2^4 + g_Z^4}{16\lambda_0^2} \right) \ln(R\mu) + R \text{ independent} \right] \end{aligned} \quad (7.7)$$

in agreement with [15].

Since the action depends on R , the degeneracy over R is broken. At this order, however, the rate can be made arbitrarily small for any choice of μ since it is a monotonic function of R . At the next order there will be terms quadratic in $\ln R\mu$. If we had those terms, we could solve for an exact, well-defined minimum of the two-loop rate, which could only give us $R_M \sim \mu^{-1}$. Thus, even at NNLO, the result would depend on an arbitrary scale μ . The full exact rate, however, is independent of the artificial scale μ . Thus, there must be an actual scale R_M for which the rate is maximal. By dimensional analysis, R_M^{-1} is almost certainly near the scale μ_0 where $\beta_\lambda(\mu_0) = 0$ [15,19]. However, to our knowledge this has not been rigorously shown. Assuming μ_0 is the correct scale, we take $R = R_M = \mu_0^{-1}$ and then $\mu = \mu_0$ to minimize the large logarithms. The conclusion is that the decay rate is given as the formula in Eq. (7.5) plus NLO corrections. The leading-order rate is so small that the NLO corrections are not even worth computing.

The take-home lesson from the analysis of the SM decay rate is that everything hinges on a single dimensionful scale μ_0 . This scale sets the size of the bubble $R = R_M \sim \mu_0^{-1}$, which in turn provides the dimensions of the decay rate $\Gamma \sim \frac{1}{R^4}$. The decay rate is exponential in λ_0^{-1} , where $\lambda_0 = \lambda(\mu_0)$. The scale μ_0 is completely undetermined by the tree-level action, which is scale invariant. It is also not determined by the NLO corrections to the decay rate: Quadratic fluctuations in the direction of changing R are not exponentially suppressed. Thus, one needs to go to at least NNLO to fix μ_0 . Looking at the effective potential at NNLO, we see that the degeneracy is broken at a scale μ_0 where $\lambda'_{\text{eff}}(h = \mu_0) = 0$, with $\lambda_{\text{eff}} \equiv \frac{4}{h^4} V_{\text{SM}}$. However, the full effective action (with derivative terms) could change this conclusion and is not known to NNLO. Of course, the simplest procedure to determine μ_0 is to maximize the leading-order rate $\Gamma \sim \exp(-\frac{8\pi^2}{|\lambda(\mu_0)|})$, as is often done [15,18,19]. The discussion here has investigated to what extent that procedure can be rigorously justified.

C. Higher-dimension operators

In the discussion above, we assumed that there was no physics beyond the Standard Model which could affect the lifetime of our vacuum. This assumption is not valid. At a minimum, there will be contributions from gravity which

come in at a scale $M_{\text{Pl}} \sim 10^{19}$ GeV, but there is also the possibility of new physics at scales well below M_{Pl} . It has been argued relatively recently by Branchina *et al.* that even contributions at M_{Pl} can destabilize the vacuum [19–22]. Their argument relies on the coincidence between the field value ϕ_0 at the center of the critical bubble (ultimately determined by where $\beta_\lambda = 0$) in the Standard Model and the Planck scale. In this section, we discuss how sensitive the SM tunneling rate may be to physics at a new scale. We confirm the Planck sensitivity, but also show how it would persist even without the coincidence between M_{Pl} and ϕ_0 .

To begin, we recall that the Standard Model is qualitatively very similar to a simple toy scalar field theory with potential $V = \frac{\lambda}{4}\phi^4$ with $\lambda < 0$. As long as we are concerned with energy scales well below the scale of the new physics, we can perform the path integral over the new particles to generate a low-energy effective action which can be expanded in derivatives. This leads us to consider a modified potential of the form

$$V = \frac{\lambda}{4}\phi^4 - \frac{1}{6\Lambda^2}\phi^6 + \frac{1}{8M^4}\phi^8 \quad (7.8)$$

where Λ and M are two parameters with dimension of mass. For $\lambda < 0$, $\Lambda > 0$ and $M > 0$, the potential has a local maximum at $\phi = 0$ and a minimum at $\phi_{\text{min}} \approx \frac{M^2}{\Lambda} + \mathcal{O}(\frac{\lambda\Lambda^3}{M^2})$.

To be clear, since we have completely integrated out the new physics, we can put this classical potential into the action to find the tunneling rate; there is no double counting since the fluctuations of new physics are integrated out in producing the potential and only the fluctuations of ϕ are used to calculate the rate. This is consistent with the discussion in Sec. VI.

While we do not have analytical solutions to the Euclidean equations of motion for this potential, we can easily find numerical solutions (see Appendix C). We find,

$$S_E[\phi_0] = \frac{162\pi^2\Lambda^2M^4(64(4\sqrt{3}-3)M^4\phi_0^2 - 27\Lambda^2(8(4\sqrt{2}-3)\lambda M^4 + 5\phi_0^4))}{(27\Lambda^2\phi_0^4 + 4M^4(27\sqrt{2}\lambda\Lambda^2 - 8\sqrt{3}\phi_0^2))^2}. \quad (7.13)$$

Minimizing with respect to ϕ_0 leads to

$$\phi_0 = 1.1 \times \frac{M^2}{\Lambda}, \quad R = 8.9 \times \frac{\Lambda^3}{M^4}, \quad S_E[\phi_G] = 224.7 \times \frac{\Lambda^4}{M^4} \quad (7.14)$$

in parametric and reasonable numerical agreement with Eq. (7.9).

numerically, that the starting point for the bounce (field value at the center) and value of the Euclidean action on the bounce are

$$\phi_0 \approx 0.85 \times \frac{M^2}{\Lambda}, \quad S_E \approx 290 \times \frac{\Lambda^4}{M^4}, \quad (7.9)$$

respectively.

1. Approximate solutions

It is perhaps informative to compare this exact numerical result to the approximate result coming from using approximate solutions, the Gaussian bubbles discussed in Sec. V C. These are bubbles of the form $\phi_G(\rho) = \phi_0 \exp(-\rho^2/R^2)$ with two parameters ϕ_0 and R , as in Eq. (5.15). The Euclidean action with this potential on these bubbles is

$$S_E[\phi_G] = \frac{1}{2}\pi^2 R^2 \phi_0^2 + \frac{\pi^2 R^4 \phi_0^4}{64} \left(\lambda - \frac{8\phi_0^2}{27\Lambda^2} + \frac{\phi_0^4}{8M^4} \right). \quad (7.10)$$

The energy of the bubbles is

$$U[\phi_G] = \frac{3\pi^{3/2} R \phi_0^2}{4\sqrt{2}} + \phi_0^4 R^3 \frac{\pi^{3/2}}{16\sqrt{2}} \left(\frac{\lambda}{\sqrt{2}} - \frac{4\phi_0^2}{9\sqrt{3}\Lambda^2} + \frac{\phi_0^4}{8M^4} \right). \quad (7.11)$$

We recall that before minimizing, we need to restrict ourselves to the $U = 0$ surface. So the zero-energy surface is

$$R^{-2} = -\lambda \frac{\phi_0^2}{12\sqrt{2}} + \frac{\phi_0^4}{27\sqrt{3}\Lambda^2} - \frac{\phi_0^6}{96M^4}. \quad (7.12)$$

This gives us the action on the zero-energy surface as a function of ϕ_0 only:

Alternatively, solving $\lambda'_{\text{eff}}(\phi_0) = 0$ gives

$$\phi_0 = \sqrt{\frac{2}{3}} \frac{M^2}{\Lambda} = 0.82 \times \frac{M^2}{\Lambda}, \quad (7.15)$$

also in good agreement with Eq. (7.9).

2. High-scale operators in the standard model

Now let us consider what happens when we add such higher dimension operators to the Standard Model. Importantly, there are two qualitatively different types of

effects. First, adding a potential V_{new} will cause a perturbative change to the Euclidean action on the Standard Model bounce, given by

$$\Delta S = \int d^4x V_{\text{new}}[h_C^R(x)] \sim \frac{1}{\Lambda^2 R^2} + \frac{1}{M^4 R^4}. \quad (7.16)$$

By dimensional analysis, this correction will be suppressed by factors of MR and ΛR as shown. Even for Planck scale operators, this may not be much suppression, since in the Standard Model $R^{-1} \sim 10^{17}$ GeV, which is not that far from M_{Pl} .

In the limit that M and Λ are taken very large, one expects that the new physics should decouple and the Standard Model bounce should be unaffected, which indeed is what happens. However, in this limit a new tunneling direction can open up; although the Standard Model bounce is unaffected, the decay rate (determined by the integration over all bounces) *can* be affected if a new direction in field space has lower Euclidean action than the Standard Model bounce. Thus, the second effect which can happen when new operators are added is that tunneling proceeds through an entirely new direction in field space.

To see the new direction emerge, we need to keep track of two corrections to the tree-level quartic potential: the Standard Model logarithms and the high-scale operators. Each of these induces a slight change to the action, $\Delta S(R)$. The high-scale operators' correction can be calculated analytically; the Standard Model logarithms must be computed numerically. These corrections pick out a dominant value of R . Depending on the scale of new physics, we might get one or two local minima (see Fig. 17).

Explicitly, taking the bounce to be the quartic bounce of size R from Eq. (5.9) (with $\lambda \rightarrow \frac{\lambda}{4}$ and denoting this solution by h_C^R), the shift in the action from the high-scale operators is

$$\begin{aligned} \Delta S_{\text{NP}}(R) &= \int \Omega_3 \rho^3 d\rho \left(-\frac{1}{6\Lambda^2} h_C^R(\rho)^6 + \frac{1}{8M^4} h_C^R(\rho)^8 \right) \\ &= \frac{64\pi^2}{\lambda^4} \left(\frac{4}{21M^4 R^4} - \frac{|\lambda|}{15\Lambda^2 R^2} \right). \end{aligned} \quad (7.17)$$

This action is minimized for

$$R_{\text{min}} = \sqrt{\frac{40}{7|\lambda|}} \frac{\Lambda}{M^2}. \quad (7.18)$$

That is, as $M \rightarrow \infty$ and $\Lambda \rightarrow \infty$ holding M/Λ fixed, tunneling is dominated by bubbles of smaller and smaller size.

Next we focus on the Standard Model corrections, which will pick out the minimum of $\lambda_{\text{eff}} = \frac{4}{h^4} V$. Of course, as we have argued, one cannot use the SM effective potential for tunneling calculations. However, there does exist a full effective action for the SM which one could use in principle, and we will assume the higher derivative terms are independent of R . Evaluating this action on bubbles of size R , one finds a function $\Delta S_{\text{SM}}(R)$ which has its minimum at R_{SM} where $R_{\text{SM}}^{-1} = \mu_0 \approx 10^{17}$ GeV. Thus, the full action is

$$S_{\text{full}} = S_{\text{SM}}^{\text{LO}} + \Delta S_{\text{SM}}(R) + \Delta S_{\text{NP}}(R). \quad (7.19)$$

We know the second term, ΔS_{SM} , has a minimum around R_{SM} , and the third term, ΔS_{NP} , has a minimum around $R \sim \frac{\Lambda}{M^2}$. So we generically expect the curve to look something like the right plot in Fig. 17.

When Λ and M are much larger than M_{Pl} , there are two local minima (see Fig. 17): the Standard Model one at R_{SM} , perturbed only slightly by the higher-dimension operators,

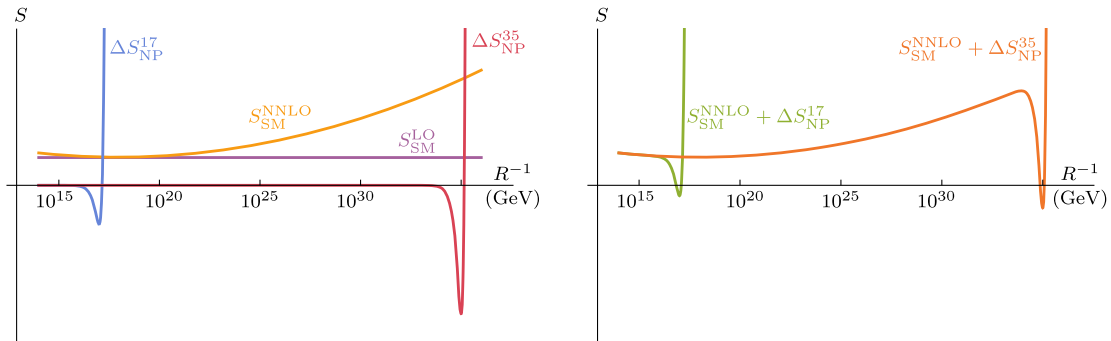


FIG. 17. Left panel: Corrections to the standard model LO action (violet) for each R , using $\mu_0 = 10^{17}$ and $\lambda(\mu_0) = -0.015$, induced by the Standard Model logs at NNLO (orange), and also the higher-dimension potential terms from Eq (7.8) with $\Lambda = 5.09 \times 10^{18}$ GeV, $M = 3.16 \times 10^{18}$ GeV (blue), and $\Lambda = 2.80 \times 10^{36}$ GeV, $M = 2.34 \times 10^{36}$ GeV (red). Right panel: Sum of Standard Model logs at NNLO and new physics. For S_{NP}^{35} there are two distinct local minimum bubbles, the SM one with $R \sim 10^{17}$ GeV and the high-scale one with $R \sim 10^{35}$ GeV, while for S_{NP}^{17} there is only one bubble. For these particular parameters for S_{NP}^{35} , the high-scale one has less total action and hence it dominates the rate.

and the higher-dimension one at $\sqrt{\frac{40}{7|\lambda|}} \frac{\Lambda}{M^2}$, perturbed only slightly by the Standard Model logs.

Since there are several bounces, the total tunneling rate is dominated by whichever has the smaller Euclidean action. The Standard Model bounce's action is given in the exponent of Eq. (7.7), and the higher-dimension operators' action is controlled by Λ and M according to Eq. (7.9). Thus, depending on the details of the higher-dimension operators, the Standard Model or the new physics bounce might dominate, no matter how high the scale Λ . This was the point illustrated by Fig. 15. Because of this possibility, the full tunneling rate is *always* sensitive to physics at arbitrary high scales. The Standard Model calculations at low energy provide a lower bound on the tunneling rate, but arbitrarily high scale operators can always make the rate faster by adding new tunneling directions in field space.

VIII. SUMMARY AND CONCLUSIONS

The first four sections of this paper attempt to provide a thorough exposition of the various ways tunneling rates are calculated in quantum mechanics and quantum field theory. While many of the methods discussed here are explained in the literature, we hoped that compiling them with explicit examples and additional commentary on subtle points rarely emphasized could be helpful. One path-integral method, which we call the direct method, was introduced in [8], and more details are given here for the first time. Later in this paper, we explored tunneling in quantum field theory and discussed the role that the effective potential plays. As an application to the Standard Model, we investigated the UV sensitivity of the tunneling rate.

One critical observation about tunneling is that two time scales must be well separated for the tunneling rate to even be well defined. First, there is what we call the sloshing time T_{slosh} , characterizing the time scale for movement within the false-vacuum well region. One must average over times of order T_{slosh} in defining the rate Γ . We call the other time scale T_{NL} . It represents the time scale for the transmitted wave function to start propagating back into the false vacuum.

In many presentations, one takes $T \rightarrow \infty$ to find the decay rate. However, as we repeatedly emphasize, this has to be done carefully. In the strict $T \rightarrow \infty$ limit, the system ends up in the true vacuum, at rest. Much of the subtlety in calculating decay rates can be traced to enforcing the double limit $T \gg T_{\text{slosh}}$ and $T \ll T_{\text{NL}}$.

In QM, one can enforce $T \ll T_{\text{NL}}$ by using radiative outgoing-wave-only boundary conditions. These boundary conditions, by definition, prevent the transmitted wave from returning. Moreover, the boundary conditions are unphysical, so the energy eigenstates of the Hermitian Hamiltonian can have complex energy eigenvalues. The imaginary part of the energy of a resonance localized in the false vacuum can be readily identified with the decay rate. This was shown in Sec. II.

The outgoing-boundary condition way of enforcing $T \ll T_{\text{NL}}$ is not readily generalizable to multidimensional systems. In the path-integral derivation, originally due to Callan and Coleman [1], this time scale is left implicit. Nevertheless, we show that if one blindly takes $T \rightarrow \infty$, the energy E_0 picked out by this method is not the quasistable false-vacuum state of interest, but rather the true vacuum (in the path integral, the dominant saddle point is not the bounce but rather the shot, a solution that stays in the true vacuum for nearly all times). To isolate the false-vacuum energy as E_0 , one must change the contour on integration to be the steepest-descent contour through the false vacuum.

We spent some time in this paper discussing various subtleties in performing the analytic continuation or contour deviation and saddle-point approximation properly. One must account for the locations of all the relevant saddle points in the complex plane, how the integration contour deforms as the continuation is done, and how the imaginary part is to be extracted so that one is using an asymptotic expansion consistently. Importantly, the relevant contour is the contour of steepest descent through the false-vacuum saddle point.

One hiccup of the potential-deformation method that we were not able to resolve is why the imaginary number computed through this method is the decay rate of the false vacuum. Unlike the outgoing-boundary conditions prescription, we found no intuitive physical narrative to connect this imaginary number to the decay rate of interest. It does seem that deforming the contour of integration to the steepest descent through the false vacuum has some connection to enforcing $T \ll T_{\text{NL}}$, which prevents the true vacuum from dominating. However, we do not know how to make this precise. Moreover, the imaginary part depends on how the analytic continuation is done. We expect that there should be a proof that this method, with some precise prescription for how the contour deformation is to be done, will always generate the decay rate of interest. It would also be great to see a proof of the universality of the famous factor of $\frac{1}{2}$ (cf. Sec. III), other than the indirect proof we provide here through the direct method. Such investigations could provide fruitful avenues for future research.

In the direct method, explored at length in Sec. III, the computation of tunneling rates is approached from the starting point of the casual propagator. Inspired by the WKB approximation, the direct method computes the rate to propagate from the false vacuum through a barrier. The condition $T \ll T_{\text{NL}}$ is imposed by assuming that once the particle exits the barrier, it will never pass back through. The end result of this method is a nonperturbative formula for the tunneling rate. It is exact, up to the exponentially small corrections that are an inherent ambiguity in the definition of the decay rate. The formula relates Γ to a ratio of two path integrals. In the saddle-point approximation to this formula, the numerator is dominated by the Euclidean bounce solution and the denominator by the static false-vacuum solution. The final prescription is in agreement

with previous results at NLO. One corollary of the direct derivation provided here is that it relates the bounce action to the actual tunneling rate thereby validating the universality of the factor of $\frac{1}{2}$.

The remainder of this paper was devoted to reviewing and expounding some important aspects of tunneling calculations in quantum field theory. We discussed how the energy functional $U[\phi]$ can be productively visualized along a tunneling direction. We then showed how one can get good intuition for tunneling and bounces from approximate analytic solutions, rather than exact numerical ones. In particular, we derived that the field at the center of the bounce $\phi_0 = \phi_B(\rho = 0)$ will generically be determined by the scale where $\partial_\phi[\frac{1}{\lambda(\mu)} V(\phi)] = 0$. This is in accordance with renormalization-group-based arguments predicated on minimizing $\Gamma \sim \exp[-\frac{1}{\lambda(\mu)}]$, but it is conceptually cleaner: It is just a shortcut to solving classical equations of motion.

One point we thought worth clarifying is the use of the effective potential in tunneling calculations. It is *not* acceptable to use an effective potential for tunneling, e.g. using it to find bounces. We also showed that the terms in the effective action with derivatives are equally important as the NLO corrections to the potential, and the higher derivative terms are generally unknown. To verify this explicitly, we computed the four-derivative terms at one loop in a scalar field theory and compared to what one gets using the full NLO bounce action. The potential terms, the four derivatives, and even higher-order terms, all contribute to the decay rate at the same order.

The last section was devoted to some comments on vacuum stability in the Standard Model. It would be good to check that the decay rate in the Standard Model is gauge invariant by an explicit calculation. Although the rate *must* be gauge invariant, and general nonperturbative proofs show that it is, there may be subtleties with the power counting that require special care when working in perturbation theory. Such was the case for the absolute stability bound [64]. As a small step in this direction, we argued that the appropriate power counting should be the usual loop expansion in \hbar , not the $\lambda \sim \hbar$ counting of the Coleman-Weinberg model. Even then, one cannot use a resummed potential or action since these mix orders in perturbation theory. Conveniently, we showed that for the effective potential at least resummation is not necessary: A fixed-order expansion using $\mu \sim 10^{17}$ GeV agrees qualitatively quite well with the full resummed potential.

Finally, we included some remarks on the UV sensitivity of the Standard Model decay rate. It has been argued that the rate is sensitive to Planck scale physics due to a coincidence between the critical bubble size in the gravity-free Standard Model, and the Planck scale. We showed that this is not true. No matter how high the Planck or UV scale is, tunneling rates will always be UV sensitive. Decoupling arguments simply do not apply to the lower bound on tunneling rates. This result is in agreement with other recent

work [18]. Without any gravitational or beyond-the-Standard-Model physics, the lifetime of our Universe appears to be around 10^{600} years. Although physics beyond 10^{18} GeV could make this lifetime shorter, it seems hard for new physics at this scale to make the lifetime longer. Thus if, for whatever reason (e.g. [15]), one can argue that our Universe must be absolutely stable, then there must be sub-Planckian physics beyond the Standard Model. Otherwise, unless we get a better handle on quantum gravity, the fate of our Universe will remain uncertain.

ACKNOWLEDGMENTS

The authors would like to thank D. Harlow, M. Marino, A. Strumia, M. Unsal, and E. Weinberg for helpful discussions. This research was supported in part by the U.S. Department of Energy, under Grant No. DE-SC0013607. A. A. is supported in part by the Stolt-Nielsen Fund for Education of the American-Scandinavian Foundation and the Norway-America Association.

APPENDIX A: COMPLEX ENERGIES AND DECAY RATES IN A SQUARE WELL

In this appendix, we compute the decay rate for the simple square-well potential shown in Fig. 18. The explicit results here may help elucidate some of the general statements from Sec. II A. Related calculations can be found, for example, in [65].

Suppose we have an initial state $|\psi\rangle$ localized in the FV region at $t = 0$. This can be a Gaussian, a delta function, or simply a constant in the FV region. At a later time, the state is $|\psi(t)\rangle = e^{-iHt}|\psi\rangle$. The probability that we find the state in region FV at a later time is given by

$$\mathcal{P}_{\text{FV}}(t) = \int_{\text{FV}} dx |\langle x|\psi(t)\rangle|^2 = \int_0^a dx |\psi(x, t)|^2. \quad (\text{A1})$$

To compute the probability, we start by decomposing the wave function into energy eigenstates. Labeling these states ϕ_p by momentum p , we can write $\psi(x, t)$,

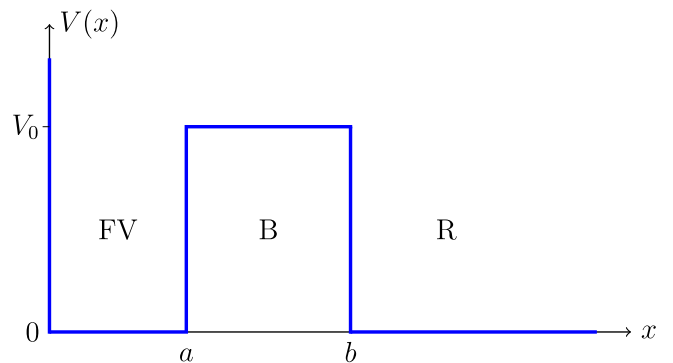


FIG. 18. Square potential well. The potential is divided into three regions: FV is the false vacuum, B is the barrier, and R is the destination region to which a state initially localized in FV will decay over time.

$$\psi(x, t) = \int_0^\infty \frac{dp}{2\pi} \int_0^\infty dy \psi(y) \phi_p^*(y) \phi_p(x) e^{-i\frac{p^2}{2m}t}, \quad (\text{A2})$$

where $\psi(y) \equiv \psi(y, t=0)$.

Since region R is infinite in extent, there are energy eigenstates for any energy $E = \frac{p^2}{2m}$. Of these, some will be

resonances. These resonances come in bands of width Γ_i around energies E_i . The resonant energies E_i are close to what the bound-state energies in the FV region would be if we disallowed tunneling, for example, with $b \rightarrow \infty$. For finite b , the resonant energies broaden across a band and have support in the region R.

More precisely, the exact energy eigenstates are

$$\phi_p(x) = \begin{cases} \phi_p^{\text{FV}}(x) = \frac{2}{N_p} \sin(px) & 0 < x < a \\ \phi_p^B(x) = \frac{1}{N_p} [A_p e^{\kappa(x-a)} + B_p e^{-\kappa(x-a)}] & a < x < b \\ \phi_p^R(x) = \frac{1}{N_p} [C_p e^{ip(x-b)} + D_p e^{-ip(x-b)}] & b < x \end{cases} \quad (\text{A3})$$

where $\kappa = \sqrt{2mV_0 - p^2}$ and

$$A_p = \sin(pa) + \frac{p}{\kappa} \cos(pa), \quad (\text{A4})$$

$$B_p = \sin(pa) - \frac{p}{\kappa} \cos(pa), \quad (\text{A5})$$

$$C_p = \frac{1}{2} \left(1 - i \frac{\kappa}{p} \right) A_p e^{W_p} + \frac{1}{2} \left(1 + i \frac{\kappa}{p} \right) B_p e^{-W_p}, \quad (\text{A6})$$

$$D_p = \frac{1}{2} \left(1 + i \frac{\kappa}{p} \right) A_p e^{W_p} + \frac{1}{2} \left(1 - i \frac{\kappa}{p} \right) B_p e^{-W_p}. \quad (\text{A7})$$

Here $W_p = \int_a^b dx \kappa = (b-a)\kappa$ is the usual WKB exponent.

The factor N_p can be computed by requiring that the states have the usual normalization

$$\int_0^\infty dx \phi_p(x) \phi_{p'}^*(x) = \delta(p - p'). \quad (\text{A8})$$

The only place such a δ function can come from is the integral over the region $b < x < \infty$. To see this, write

$$\begin{aligned} & \int_0^\infty dx \phi_p(x) \phi_{p'}^*(x) \\ &= \int_b^\infty dx \phi_p^R(x) \phi_{p'}^{R*}(x) + \int_0^b dx \phi_p(x) \phi_{p'}^*(x) \end{aligned} \quad (\text{A9})$$

$$= \pi \frac{C_p C_{p'}^* + D_p D_{p'}^*}{N_p N_{p'}^*} \delta(p - p'). \quad (\text{A10})$$

The second integral, from 0 to b , has exactly vanished. Comparing with Eq. (A8) and noting that $C_p = D_p^*$ for real p , we can write $|N_p|^2$ as an analytic function of p :

$$|N_p|^2 = 2\pi C_p D_p. \quad (\text{A11})$$

We now see that up to an overall phase, $\phi_p^R(x) = \cos(px)$ for any p . That is, all the wave functions are order 1 in the region R. In the FV region, the wave functions vary in size as $\frac{1}{|N_p|}$. Equations (A6), (A7), and (A11) imply that $N = \sqrt{2\pi} C \sim A e^W + B e^{-W}$. For $W \gg 1$, where the WKB approximation is supposed to work, $N \sim A e^W$. Thus, generically, $\phi_p^{\text{FV}} \sim e^{-W}$. So for most values of p , the wave function has support almost entirely outside the well and is exponentially suppressed in the well. The only time it can have reasonable support in the well is when $A \lesssim e^{-2W}$; indeed for $A \lesssim e^{-2W}$, we see $N \lesssim e^{-W}$ and $\phi_p^{\text{FV}} \sim e^W$.

The momenta for which $|N_p|^2$ is minimized are the resonance momenta. They are exponentially close [within $\mathcal{O}(e^{-2W})$] to the zeros of A_p . The zeros of A_p correspond to the bound states in the limit $b \rightarrow \infty$ or $W \rightarrow \infty$. This can be seen by considering the condition for wave-function normalization in this limit; it is precisely $A_p = 0$ that stops the wave function's growth as x increases, and discretizes the spectrum. For finite b , tunneling through the barrier shifts the bound-state energies by $\mathcal{O}(e^{-2W})$, and they become the resonant energies.

1. Relating the probability to the pole

To calculate the probability $P_{\text{FV}}(t)$, let us assume for simplicity that the initial wave function $\psi(x) \equiv \psi(x, t=0)$ only has significant overlap with modes whose energies are close to E_0 , where E_0 is the smallest real energy for which $|N(E)|^2$ has a local minimum. Here, we are interchanging the momentum label p with an energy label E , where $p^2 = 2mE$. So we will write $C(E)$, $D(E)$, etc., rather than C_p , D_p , etc.

To compute the probability, it is helpful to consider imaginary energies. To do so, we first need to analytically continue $|N(E)|^2$, which can be done by writing it in the

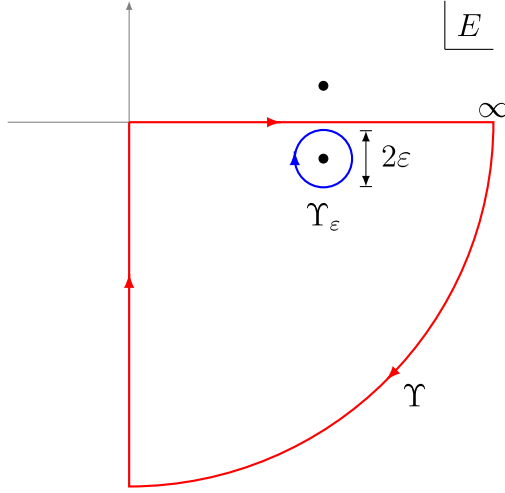


FIG. 19. The contour Υ used in Eq. (A14) enclosing the pole $E = E_0 - \frac{i}{2}\Gamma_0$, and the smaller contour Υ_ε allowing the use of the Laurent series.

form of Eq. (A11): $|N(E)|^2 = 2\pi C(E)D(E)$. This analytic function has zeros in the complex plane, and the zeros come in pairs. Indeed, from looking at the form of C and D in Eq. (A6), we see that if $E = a + ib$ is a zero of C , then $E^* = a - ib$ will be a zero of D . From the form of ϕ^R in Eq. (A3), we see that $C = 0$ corresponds to incoming boundary conditions and $D = 0$ to outgoing boundary

conditions. The first pair of zeros for $|N|^2 = 2\pi CD$ are at $E = E_0 \pm \frac{i}{2}\Gamma_0$. As we will confirm, $\frac{\Gamma_0}{E_0} \sim e^{-2W} \ll 1$ as in the WKB approximation.

Plugging Eq. (A3) into Eq. (A2) we find

$$\psi(x, t) = \int_0^a dy \psi(y) \int_0^\infty dE \left(\frac{1}{2\pi} \sqrt{\frac{m}{2E}} \right) \times \frac{\sin(\sqrt{2mEx}) \sin(\sqrt{2mEy})}{2\pi C(E)D(E)} e^{-iEt} \quad (\text{A12})$$

where the spatial integral is over $y \in (0, a)$ since $\psi(y, t=0)$ only has support in the FV region. This is convergent when E has a negative imaginary part. So we can deform the contour to write

$$\psi(x, t) = \int_0^a dy \psi(y) [F(x, y, t) + G(x, y, t)] \quad (\text{A13})$$

where

$$F(x, y, t) = \oint_\Upsilon dE \left(\frac{1}{2\pi} \sqrt{\frac{m}{2E}} \right) \times \frac{\sin(x\sqrt{2mE}) \sin(y\sqrt{2mE})}{2\pi C(E)D(E)} e^{-iEt} \quad (\text{A14})$$

and

$$\begin{aligned} G(x, y, t) &= \int_0^{-i\infty} dE \left(\frac{1}{2\pi} \sqrt{\frac{m}{2E}} \right) \frac{\sin(x\sqrt{2mE}) \sin(y\sqrt{2mE})}{2\pi C(E)D(E)} e^{-iEt} \\ &= -i \int_0^\infty d\mathcal{E} \left(\frac{1}{2\pi} \sqrt{\frac{m}{-i2\mathcal{E}}} \right) \frac{\sin(x\sqrt{-i2m\mathcal{E}}) \sin(y\sqrt{-i2m\mathcal{E}})}{2\pi C(-i\mathcal{E})D(-i\mathcal{E})} e^{-\mathcal{E}t}. \end{aligned} \quad (\text{A15})$$

The contour Υ for the integral in $F(x, y, z)$ is shown in Fig. 19.

The $F(x, y, t)$ integral can be calculated by replacing the contour Υ with Υ_ε , a circle of radius ε around the pole. This lets us expand $C(E)D(E)$ around its zero and use its lowest order term. That is, we can write $C(E)D(E) = C(E_0 - \frac{i}{2}\Gamma_0)D'(E_0 - \frac{i}{2}\Gamma_0)(E - E_0 + \frac{i}{2}\Gamma_0) + \dots$ and use the residue theorem to give

$$\begin{aligned} F(x, y, t) &= 2\pi i \left(\frac{1}{2\pi} \sqrt{\frac{m}{2(E_0 - \frac{i}{2}\Gamma_0)}} \right) \frac{\sin\left(x\sqrt{2m(E_0 - \frac{i}{2}\Gamma_0)}\right) \sin\left(y\sqrt{2m(E_0 - \frac{i}{2}\Gamma_0)}\right)}{2\pi C(E_0 - \frac{i}{2}\Gamma_0)D'(E_0 - \frac{i}{2}\Gamma_0)} e^{-i(E_0 - \frac{i}{2}\Gamma_0)t} \\ &= \frac{i}{2\pi} \sqrt{\frac{m}{2E_0}} \frac{\sin(x\sqrt{2mE_0}) \sin(y\sqrt{2mE_0})}{C(E_0)D'(E_0)} e^{-iE_0 t} e^{-\frac{1}{2}\Gamma_0 t} \left[1 + \mathcal{O}\left(\frac{\Gamma_0}{E_0}\right) \right] \end{aligned} \quad (\text{A16})$$

where in the last line we have used $\frac{\Gamma_0}{E_0} \ll 1$.

To evaluate $G(x, y, t)$, we begin by replacing $\mathcal{E} \rightarrow E_0\alpha$ in Eq. (A15) where α is dimensionless. Then

$$G(x, y, t) = -iE_0 \left(\frac{1}{2\pi} \sqrt{\frac{m}{-i2E_0}} \right) \int_0^\infty d\alpha \frac{1}{\sqrt{\alpha}} \frac{\sin(\sqrt{-i2mE_0}x\sqrt{\alpha}) \sin(\sqrt{-i2mE_0}y\sqrt{\alpha})}{2\pi C(-iE_0\alpha)D(-iE_0\alpha)} e^{-\alpha(E_0 t)}. \quad (\text{A17})$$

Now let us assume that $t \gg \frac{1}{E_0} \sim T_{\text{slosh}}$. This is one of the required conditions for having a well-defined decay rate. Then $E_0 t \gg 1$ and, due to the exponential factor, only very small values of α contribute. Thus we can expand the prefactor for small α , giving

$$\begin{aligned}
G(x, y, t) &= -iE_0 \left(\frac{1}{2\pi} \sqrt{\frac{m}{-i2E_0}} \right) \int_0^\infty d\alpha \frac{1}{\sqrt{\alpha}} \frac{-i2mE_0 x y \alpha}{|N(0)|^2} [1 + \mathcal{O}(\alpha)] e^{-\alpha(E_0 t)} \\
&= -(1+i)(E_0 m)^{\frac{3}{2}} \frac{xy}{4\sqrt{\pi}|N(0)|^2} (E_0 t)^{-\frac{3}{2}} [1 + \mathcal{O}((E_0 t)^{-1})].
\end{aligned} \tag{A18}$$

Now note that $F \sim \exp(-\frac{1}{2}\Gamma_0 t)$ while $G \sim (E_0 t)^{-3/2}$. Thus if t is not too large, $t \lesssim \Gamma_0^{-1} \sim T_{\text{NL}}$, $F \gg G$ and

$$\psi(x, t) \approx \left[\int_0^a dy \psi(y) \sin(\sqrt{2mE_0} y) \right] \frac{1}{4\pi^2} \sqrt{\frac{m}{2E_0}} \frac{\sin(\sqrt{2mE_0} x)}{C(E_0)D'(E_0)} e^{-iE_0 t} e^{-\frac{1}{2}\Gamma_0 t} \tag{A19}$$

where \approx means that terms $\mathcal{O}((E_0 t)^{-\frac{3}{2}})$ and higher order in $\frac{\Gamma_0}{E_0}$ have been dropped. The probability in Eq. (A1) therefore takes the form

$$P_{\text{FV}}(t) \approx \text{const} \times e^{-\Gamma_0 t} \tag{A20}$$

$E_0^{-1} \ll t \ll \Gamma_0^{-1}$

and thus Γ_0 is indeed the rate.

In other words, we have established a direct connection between the complex zeros of $D(E)$ and the decay rate. From Eq. (A3), the complex zeros of $D(E)$ correspond to outgoing-only plane waves in region R. In this way, the connection between Gamow-Siegert outgoing boundary conditions and decay rates is made precise (at least in this example).

Note that the assumption $t < T_{\text{NL}} \sim \Gamma_0^{-1}$ was essential. At very late times $t \gg \Gamma_0^{-1}$, then $G \gg F$. In this limit,

$$P_{\text{FV}}(t) \approx \text{const} \times \frac{1}{(E_0 t)^3}. \tag{A21}$$

$E_0^{-1} \ll \Gamma_0^{-1} \ll t$

This is the nonlinear behavior. In this regime there is not a well-defined decay rate.

In this calculation we assumed we were dealing with an initial wave function dominated by energy eigenfunctions close to E_0 . If this were not the case, we would have other resonances to worry about, each corresponding to poles of the form $E = E_n - \frac{i}{2}\Gamma_n$. The above analysis follows through in the same way, and the result is a sum of the form $\sum_n a_n e^{-\Gamma_n t}$ for some a_n . After enough time, only the dominant term with $P \sim e^{-\Gamma_0 t}$ will remain, while the others decay away.

2. Explicit computation of Γ

We have shown that the decay rate Γ is determined by the imaginary part of the energy for which $D(E)$ vanishes. Now let us calculate Γ explicitly.

First, we observe that the complex zeros are exponentially close to the resonant energies. These resonant energies $E_R = \frac{p_R^2}{2m}$ are the zeros of $A(E)$ or equivalently of A_p . Setting $A_p = 0$ implies

$$\sin(p_R a) = -\frac{p_R}{\sqrt{p_R^2 + \kappa_R^2}}, \quad \cos(p_R a) = \frac{\kappa_R}{\sqrt{p_R^2 + \kappa_R^2}}, \tag{A22}$$

where $\kappa_R \equiv \sqrt{2mV_0 - p_R^2}$. We can then find the complex zeros of D_p by expanding perturbatively in

$$\delta \equiv e^{-W_p} = e^{-\kappa_R(b-a)}. \tag{A23}$$

Both the real and the imaginary part of the complex energies should differ from the resonant energies by amounts of order δ^2 .

Next, we write $p = p_R + \delta^2 p_C + \mathcal{O}(\delta^4)$ and expand

$$\begin{aligned}
D_p &= \frac{1}{2} \left(1 + i \frac{\kappa}{p} \right) \left[\sin(pa) + \frac{p}{\kappa} \cos(pa) \right] e^{W_p} \\
&\quad + \frac{1}{2} \left(1 - i \frac{\kappa}{p} \right) \left[\sin(pa) - \frac{p}{\kappa} \cos(pa) \right] e^{-W_p}
\end{aligned} \tag{A24}$$

to order δ . Setting $D_p = 0$ and using Eq. (A22) to simplify the answer, we find

$$p_C = \frac{2p_R \kappa_R^2}{(p_R + i\kappa_R)^2 (1 + a\kappa_R)} \tag{A25}$$

and therefore

$$\Gamma = -2\text{Im} \frac{p^2}{2m} = \frac{8p_R^3 \kappa_R^3}{m(1 + a\kappa_R)(p_R^2 + \kappa_R^2)^2} e^{-2W}. \tag{A26}$$

Note that unlike Eq. (2.2) the prefactor does depend on the height of the barrier, V_0 .

It is perhaps informative to compare this calculation to the result of Eq. (2.16). Note that Eq. (2.16) is independent of the normalization, so we can set $N_p = 1$ in Eq. (A3). Then we find

$$|\phi_E(b)|^2 = \frac{16p_R^2 \kappa_R^2}{(p_R^2 + \kappa_R^2)^2} \delta^2, \tag{A27}$$

$$\int_0^a dx |\phi_E(x)|^2 = 2a + \frac{2\kappa_R}{p_R^2 + \kappa_R^2} + \mathcal{O}(\delta^2), \quad (\text{A28})$$

and

$$\int_a^b dx |\phi_E(x)|^2 = \frac{1}{p_R^2 + \kappa_R^2} \frac{2p_R^2}{\kappa_R} + \mathcal{O}(\delta^2), \quad (\text{A29})$$

so

$$\Gamma = \frac{p_b}{m} \frac{|\phi_E(b)|^2}{\int_0^b dx |\phi_E(x)|^2} = \frac{8p_R^3 \kappa_R^3}{m(1 + a\kappa_R)(p_R^2 + \kappa_R^2)^2} e^{-2W} \quad (\text{A30})$$

in agreement with Eq. (A26).

On the other hand, if we use the real-energy eigenstate in Eq. (2.16), we would get

$$\Gamma_{\text{real}} = \frac{p_b}{m} \frac{|\phi_E(b)|^2}{\int_0^b dx |\phi_E(x)|^2} = \frac{2p_R^3 \kappa_R}{m(1 + a\kappa_R)(p_R^2 + \kappa_R^2)} e^{-2W}. \quad (\text{A31})$$

Taking the ratio we find

$$\frac{\Gamma_{\text{real}}}{\Gamma} = \frac{1}{4} + \frac{p_R^2}{4\kappa_R^2}. \quad (\text{A32})$$

Thus using the real-energy eigenstates gets the right rate to a factor of order 1. The factor of 4 in the large κ_R (large V_0) limit can be traced to the mode that grows inside the barrier from a to b . This mode is exactly zero for the real-energy state.

APPENDIX B: CHANGING TO COLLECTIVE COORDINATES

The Gaussian integral $\int d\xi_0 \exp(-\lambda_0 \xi_0^2)$ is divergent when λ_0 is zero. This happens for the zero mode corresponding to translation invariance around a bounce. In order to regulate this divergence, the standard procedure is to trade ξ_0 for a collective coordinate τ_0 so that the integral over τ_0 simply gives a factor of \mathcal{T} which can then be divided out to get a decay rate [1,5,6]. In trading ξ_0 for τ_0 , one also must adjust the other modes ξ_i to a new orthonormal basis ζ_i . The change of variables from $\{\xi_i\}$ to $\{\tau_0, \zeta_i\}$ results in a Jacobian factor $J(\tau_0, \zeta_i)$. This appendix addresses some subtleties in the change of variables that we have not seen in decay-rate literature but was understood in pioneering works on collective coordinates [66].

In the evaluation of the saddle-point approximation around a bounce or instanton denoted \bar{x} , one naturally parametrizes paths by the eigenfunctions x_n of S_E'' with eigenvalues λ_n :

$$x^{\xi_0, \xi_1, \dots}(\tau) = \bar{x}(\tau) + \sum_{n=0}^{\infty} \xi_n x_n(\tau). \quad (\text{B1})$$

One of these modes, $x_0 = \partial_\tau \bar{x}$, has an eigenvalue of exactly zero: $\lambda_0 = 0$. This zero mode corresponds to an infinitesimal shift. Large shifts in τ are an approximate symmetry of the path integral using Dirichlet [1] boundary conditions as an exact symmetry with periodic boundary conditions [5,6]. In order to make the symmetry manifest, it is helpful to have one of the coordinates parametrizing paths be the shift by τ_0 rather than the addition of an ξ_0 amount of the first derivative.

We might write

$$x^{\tau_0, \xi_0, \xi_1, \dots}(\tau) = \bar{x}(\tau - \tau_0) + \sum_{i=1}^{\infty} \xi_i x_i(\tau) \quad (\text{B2})$$

as in [6]. But unfortunately this parametrization is not complete. To see that, integrate both sides of Eq. (B2) against $x_0(t)$. Using Eq. (B1) for the left-hand side, we find $\xi_0(\tau_0) = \int d\tau x_0(\tau) \bar{x}(\tau - \tau_0)$, which is bounded. In contrast, the integration over ξ_0 should go from $-\infty$ to ∞ to parametrize all paths. In particular, the parametrization in Eq. (B2) does not cover exactly the large- ξ_0 fluctuations which caused the problem with the Gaussian integrations in the first place.

A better parametrization is

$$x^{\tau_0, \zeta_1, \zeta_2, \dots}(\tau) = \bar{x}(\tau - \tau_0) + \sum_{n=1}^{\infty} \zeta_n x_n(\tau - \tau_0). \quad (\text{B3})$$

Here we have used ζ_n instead of ξ_n for $n > 0$ since the coordinates for all the modes generically change when we change variables. This parametrization is complete.

Next, we calculate the Jacobian $J(\tau_0, \zeta)$ between the parametrizations Eqs. (B1) and (B3). There are two subtleties in this calculation that are often overlooked (e.g. in [3]):

1. Because the ζ_n are not the same as the ξ_n , this Jacobian is really the determinant of a nontrivial infinite-dimensional matrix; it is not simply equal to $d\xi_0/d\tau_0$.
2. Because $\bar{x}(\tau)$ breaks time-translation symmetry, one must show that $J(\tau_0, \zeta)$ is independent of τ . For example, in [3], only $J(0, \zeta)$ is calculated and assumed equal to $J(\tau_0, \zeta)$ [cf. Eqs. (17.103) and (17.108)].⁹

To calculate the Jacobian, we write the ξ_n as a function of ζ_m and τ_0 ,

⁹Actually, in the direct path-integral method, described in Sec. IV, only $J(0, 0)$ is needed because of the δ function in the path integral. In the conventional potential-deformation method, the full $J(\tau_0, \zeta)$ is needed.

$$\begin{aligned}
\xi_n &= \int [x(\tau) - \bar{x}(\tau)] x_n(\tau) d\tau \\
&= \int \left[\bar{x}(\tau - \tau_0) - \bar{x}(\tau) + \sum_{m=1}^{\infty} \zeta_m x_m(\tau - \tau_0) \right] x_n(\tau) d\tau
\end{aligned} \tag{B4}$$

which means that

$$\begin{aligned}
\frac{\partial \xi_n}{\partial \zeta_m} &= \int x_n(\tau) x_m(\tau - \tau_0) d\tau \\
\frac{\partial \xi_n}{\partial \tau_0} &= \int \left[-\dot{\bar{x}}(\tau - \tau_0) - \sum_{m=1}^{\infty} \zeta_m \dot{x}_m(\tau - \tau_0) \right] x_n(\tau) d\tau \\
&= -\sqrt{\frac{S_E[\bar{x}]}{m}} \int x_n(\tau) x_0(\tau - \tau_0) d\tau \\
&\quad - \sum_{m=1}^{\infty} \zeta_m \int \dot{x}_m(\tau - \tau_0) x_n(\tau) d\tau.
\end{aligned} \tag{B5}$$

To proceed, it is useful to define the orthogonal matrix U :

$$U_{nm}(\tau_0) \equiv \int d\tau x_n(\tau) x_m(\tau - \tau_0). \tag{B6}$$

Note that U is orthogonal because both $\{x_i(\tau)\}$ and $\{x_i(\tau - \tau_0)\}$ are complete bases. The derivative matrix is:

$$\left(\begin{array}{c|ccc} \vdots & \vdots & & \\ -\sqrt{\frac{S_E}{m}} U_{n0} + \sum \zeta_m \dot{U}_{nm} & U_{n1} & U_{n2} & \cdots \\ \vdots & \vdots & & \end{array} \right). \tag{B7}$$

The determinant is linear in the first column, so the determinant is

$$J(\tau_0, \zeta) = \left| -\sqrt{\frac{S_E}{m}} \det U + \det \left(\begin{array}{c|ccc} \vdots & \vdots & & \\ v_n & U_{n1} & U_{n2} & \cdots \\ \vdots & \vdots & & \end{array} \right) \right| \tag{B8}$$

where the vector v is defined by $v_n \equiv \sum \zeta_m \dot{U}_{nm}$. Since U is orthogonal, we can decompose v in terms of the columns of U ; $v_n = \sum_k c_k U_{nk}$. Then the second term in Eq. (B8) is a linear combination of determinants of U with the zeroth column replaced by the k th column. This determinant is simply 0 if $k \neq 0$ and $\det U$ if $k = 0$. Since U is orthogonal, $\det U = 1$, so we have

$$J(\tau_0, \zeta) = \left| -\sqrt{\frac{S_E}{m}} + c_0 \right|. \tag{B9}$$

The coefficient c_0 is simply the $k = 0$ component of the vector v_n decomposed into the columns of U :

$$c_0 = \sum_{n=0}^{\infty} v_n U_{n0} = \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} \zeta_m \dot{U}_{nm} U_{n0} = - \sum_{m=1}^{\infty} \zeta_m r_m \tag{B10}$$

where

$$r_m \equiv \int d\tau \dot{x}_m(\tau) x_0(\tau). \tag{B11}$$

So altogether the Jacobian is exactly

$$J(\tau_0, \zeta) = \sqrt{\frac{S_E}{m}} + \sum_{m=1}^{\infty} \zeta_m r_m \tag{B12}$$

and we see that it is indeed independent of τ_0 .

APPENDIX C: FINDING NUMERICAL BOUNCE SOLUTIONS

In this appendix we discuss how to numerically find bounce solutions. We want to solve Eq. (5.6):

$$\partial_\rho^2 \phi + \frac{3}{\rho} \partial_\rho \phi - V'[\phi] = 0 \tag{C1}$$

with boundary conditions $\phi'(0) = 0$ and $\phi(\infty) = 0$. Equivalently, we want to find an initial condition $\phi(0) = \phi_0$ for which the field rolls down the potential $-V(\phi)$ ending at the origin $\phi = 0$ at asymptotically late times. The usual shooting method suggests we try various values of ϕ_0 until we find one initial condition ϕ_0^+ for which the evolution overshoots [ends up with $\phi(\rho) < 0$ for some ρ] and one initial condition ϕ_0^- for which the evolution undershoots [$\phi(\rho) > 0$ for all ρ]. Then we know the solution is between ϕ_0^+ and ϕ_0^- , so we simply have to refine this interval until the desired precision is reached.

One difficulty with the shooting method described above is that the $\frac{3}{\rho}$ coefficient in the differential equation makes the point $\rho = 0$ singular. Thus, when numerically solving the equation, one has to start at some small $\rho_0 > 0$, say $\rho_0 = 10^{-5}$. However, taking $\phi(\rho_0) = \phi_0$ and $\phi'(\rho_0) = 0$ as boundary conditions can be dangerous. These conditions imply that ϕ has rolled from some $\phi(0)$ to end up at ϕ_0 at rest when $\rho = \rho_0$. But how does ϕ come to rest at ρ_0 ? This is only possible if it rolls *up* the potential to get to ϕ_0 and then turns around to roll back. Clearly, such a solution is not what we were looking for and will depend on ρ_0 . Often the effect of starting at ρ_0 is negligible since the rolling

starts off slow due to the friction term. However, for improved convergence, or for situations like searching for multiple bounces in which high precision is necessary, it can be helpful to reduce the ρ_0 dependence.

This difficulty can be overcome by expanding the potential around ϕ_0 :

$$V(\phi) \approx V_{\text{lin}} \equiv V(\phi_0) + (\phi - \phi_0)V'(\phi_0). \quad (\text{C2})$$

Using $V_{\text{lin}}(\phi)$ in Eq. (C1) leads to an analytic solution

$$\phi_{\text{lin}}(\rho) = \phi_0 + \frac{1}{8}\rho^2 V'(\phi_0). \quad (\text{C3})$$

So if $\phi(0) = \phi_0$ with $\phi'(0) = 0$, then

$$\phi(\rho_0) = \phi_0 + \frac{1}{8}\rho_0^2 V'(\phi_0) \quad \text{and} \quad \phi'(\rho_0) = \frac{1}{4}\rho_0 V'(\phi_0). \quad (\text{C4})$$

Using these boundary conditions allows for an efficient numerical solution to the differential equation and a fast convergence towards the bounce. The solutions computed this way are very insensitive to ρ_0 .

APPENDIX D: HIGHER DERIVATIVE CORRECTIONS

The effective action is constructed so that when used classically (at tree level) it reproduces the quantum physics (all loop order) of a classical action. Unfortunately, it is difficult, if not impossible, to calculate the effective action exactly even at one loop. Diagrammatically the effective action can be computed by summing over 1PI graphs with any number of external legs with any momenta running through them. Even at one loop, there are an infinite number of relevant graphs, so computing the effective action exactly is intractable.

Fortunately, a derivative expansion of the effective action is calculable. The 1PI effective action of a scalar field ϕ at up to four-derivative order can be written as

$$S_{\text{eff}}[\phi] = \int d^4x \left[-V_{\text{eff}}(\phi) + \frac{1}{2}(\partial_\mu \phi)^2 Z_2(\phi) + \frac{1}{2}(\Box \phi)^2 Z_4(\phi) + \frac{1}{2}(\partial_\mu \phi)^2 \Box \phi \rho(\phi) + \frac{1}{8}(\partial_\mu \phi)^2 (\partial_\nu \phi)^2 \Omega(\phi) + \mathcal{O}(\partial^6) \right]. \quad (\text{D1})$$

Using Lorentz invariance and integration by parts, we have reduced the action to depending on only five independent functions: $V_{\text{eff}}(\phi)$, $Z_2(\phi)$, $Z_4(\phi)$, $\rho(\phi)$, and $\Omega(\phi)$.

Note that V_{eff} is the well-known effective potential. To compute it, we expand the Lagrangian around a constant background field, $\mathcal{L}(\phi + \tilde{\phi})$, and calculate the vacuum diagrams where ϕ propagates and $\tilde{\phi}$ is fixed. (Throughout this appendix $\tilde{\phi}$ will represent a constant field.) To determine the other four functions, one might think to calculate the 1PI vertices using S_{eff} and \mathcal{L} , and compare, but this is not so straightforward. The problem lies in the fact that Z_2 , Z_4 , ρ , and Ω are nonlocal functions of ϕ , with terms like $\ln \frac{\phi}{\mu}$ or $\frac{1}{\phi^3}$ in them. One cannot derive Feynman rules for such terms as one does for a local Lagrangian.

To proceed, we note that the effective potential is computed by expanding around a constant background field $\tilde{\phi}$, but with momentum dependence in ϕ . In background-field calculations, the external lines, with or without momentum, are always ϕ . To find S_{eff} , we simply compute the same thing. We expand $S_{\text{eff}}[\phi + \tilde{\phi}]$ for constant $\tilde{\phi}$ and compute diagrams with external ϕ legs, with or without momentum in them. The difference between the calculation using S_{eff} and using \mathcal{L} is that with the effective action, only tree-level graphs are ever evaluated.

Let us compute with the effective action first. We take $\phi \rightarrow \phi + \tilde{\phi}$ and series expand each function around $\tilde{\phi} = 0$. Let us write $Z_i(\phi + \tilde{\phi}) = \sum_{n=0}^{\infty} \frac{1}{n!} Z_i^{(n)}(\tilde{\phi}) \phi^n$ and so on, where for each function $f^{(n)}(\tilde{\phi}) \equiv \frac{d^n}{d(\tilde{\phi})^n} f(\tilde{\phi})$. Then,

$$S_{\text{eff}}[\phi + \tilde{\phi}] = \int d^4x \sum_{n=0}^{\infty} \frac{1}{n!} \phi^n \left[-V_{\text{eff}}^{(n)}(\tilde{\phi}) + \frac{1}{2} (\partial_{\mu} \phi)^2 Z_2^{(n)}(\tilde{\phi}) + \frac{1}{2} (\Box \phi)^2 Z_4^{(n)}(\tilde{\phi}) \right. \\ \left. + \frac{1}{2} (\partial_{\mu} \phi)^2 \Box \phi \rho^{(n)}(\tilde{\phi}) + \frac{1}{8} (\partial_{\mu} \phi)^2 (\partial_{\nu} \phi)^2 \Omega^{(n)}(\tilde{\phi}) + \mathcal{O}(\partial^6) \right]. \quad (\text{D2})$$

The expanded Lagrangian is now local in ϕ , so we can easily compute Feynman diagrams that have external ϕ lines with it. We get

$$\frac{\rightarrow}{p_1} \text{---} \text{---} \text{---} \left(\text{circle} \right) \text{---} \text{---} \text{---} \frac{\leftarrow}{p_2} = -i \frac{\partial^2}{\partial \tilde{\phi}^2} V_{\text{eff}}(\tilde{\phi}) - ip_1 \cdot p_2 Z_2(\tilde{\phi}) + ip_1^2 p_2^2 Z_4(\tilde{\phi}) + \mathcal{O}(p^6) \quad (\text{D3})$$

$$\begin{aligned}
\text{Diagram: } \text{circle with } p_1 \text{ (right), } p_2 \text{ (top-left), } p_3 \text{ (bottom-left)} &= -i \frac{\partial^3}{\partial \tilde{\phi}^3} V_{\text{eff}}(\tilde{\phi}) - i \left(\sum_{i>j} p_i \cdot p_j \right) \frac{\partial}{\partial \tilde{\phi}} Z_2(\tilde{\phi}) + i \left(\sum_{i>j} p_i^2 p_j^2 \right) \frac{\partial}{\partial \tilde{\phi}} Z_4(\tilde{\phi}) \\
&+ i \left(\sum_{\substack{i>j \\ k \neq i,j}} p_i \cdot p_j p_k^2 \right) \rho(\tilde{\phi}) + \mathcal{O}(p^6)
\end{aligned} \tag{D4}$$

$$\begin{aligned}
\text{Diagram: } \text{circle with } p_1 \text{ (right), } p_2 \text{ (top-left), } p_3 \text{ (bottom-left), } p_4 \text{ (bottom-right)} &= -i \frac{\partial^4}{\partial \tilde{\phi}^4} V_{\text{eff}}(\tilde{\phi}) - i \left(\sum_{i>j} p_i \cdot p_j \right) \frac{\partial^2}{\partial \tilde{\phi}^2} Z_2(\tilde{\phi}) + i \left(\sum_{i>j} p_i^2 p_j^2 \right) \frac{\partial^2}{\partial \tilde{\phi}^2} Z_4(\tilde{\phi}) \\
&+ i \left(\sum_{\substack{i>j \\ k \neq i,j}} p_i \cdot p_j p_k^2 \right) \frac{\partial}{\partial \tilde{\phi}} \rho(\tilde{\phi}) + i \left(\sum_{\substack{i>j, k>l \\ \text{all different}}} p_i \cdot p_j p_k \cdot p_l \right) \Omega(\tilde{\phi}) + \mathcal{O}(p^6)
\end{aligned} \tag{D5}$$

This process can be continued to higher derivative terms if desired.

Now we compute the same n -point functions using the classical action, at one-loop level, also with an external ϕ field. For the example of massless ϕ^4 theory, the Lagrangian density is

$$\mathcal{L} = -\frac{1}{2} \phi \square \phi - \lambda \phi^4. \tag{D6}$$

Following the process outlined above, we expand this with $\phi \rightarrow \phi + \tilde{\phi}$, resulting in

$$\mathcal{L}_{\text{Expanded}} = -\lambda \tilde{\phi}^4 - 4\lambda \tilde{\phi}^3 \phi - \frac{1}{2} \phi [\square + 12\lambda \tilde{\phi}^2] \phi - 4\lambda \tilde{\phi} \phi^3 - \lambda \phi^4. \tag{D7}$$

We can safely drop $-\lambda \tilde{\phi}^4 - 4\lambda \tilde{\phi}^3 \phi$ because these terms cannot contribute to the diagrams we are calculating. Thus for our purposes, we have

$$\mathcal{L}_{\text{Expanded}} = -\frac{1}{2} \phi [\square + m^2] \phi - 4\lambda \tilde{\phi} \phi^3 - \lambda \phi^4 \tag{D8}$$

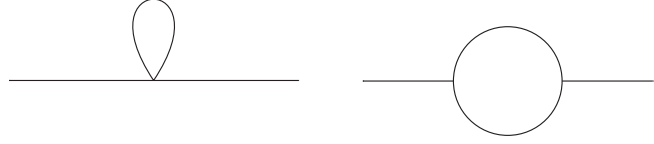
with $m^2 = 12\lambda \tilde{\phi}^2$. The Feynman rules are

$$D(k) = \text{---} = \frac{i}{k^2 - m^2}, \quad \text{---} \text{---} \text{---} = -i24\lambda \tilde{\phi}, \quad \text{---} \text{---} \text{---} = -i24\lambda \tag{D9}$$

The zero-point diagrams give the well-known one-loop effective potential [10]:

$$V_{\text{eff}} = \lambda \phi^4 - \frac{i}{2} \hbar \int \frac{d^4 k}{(2\pi)^4} \ln \left(1 - \frac{m^2}{k^2} \right) + \dots = \lambda \phi^4 + \frac{9\hbar}{4\pi^2} \lambda^2 \phi^4 \left(\ln \frac{12\lambda \phi^2}{\mu^2} - \frac{3}{2} \right) + \mathcal{O}(\hbar^2). \tag{D10}$$

There are two one-loop diagrams for the two-point function


(D11)

which, along with the tree-level piece, is given by the amplitude

$$\begin{aligned}
 \frac{\vec{p}_1}{p_1} \text{---} \text{---} \frac{\leftarrow{p_2}}{p_2} &= -i(m^2 - p_1^2) + \frac{(24\lambda\phi)^2}{2} \hbar \int \frac{d^4 k}{(2\pi)^4} \frac{1}{(k^2 - m^2)((k - p_1)^2 - m^2)} + 12\lambda\hbar \int \frac{d^4 k}{(2\pi)^4} \frac{1}{k^2 - m^2} \\
 &= -i(m^2 - p_1^2) + i \frac{\hbar}{16\pi^2} \frac{(24\lambda\phi)^2}{2} \left[-\ln(m^2) + p_1^2 \frac{1}{6m^2} + (p_1^2)^2 \frac{1}{60(m^2)^2} + \dots \right] \\
 &\quad + i \frac{\hbar}{16\pi^2} 12\lambda m^2 [1 - \ln m^2]
 \end{aligned}
 \tag{D12}$$

where in the last line, we have evaluated the integral expanded in momenta. We find that the momentum-free piece gives $-i \frac{\partial^2 V_{\text{eff}}}{\partial \phi^2}$ as expected. Using $p_1 = -p_2$, we can rewrite this in the form of Eq. (D3),

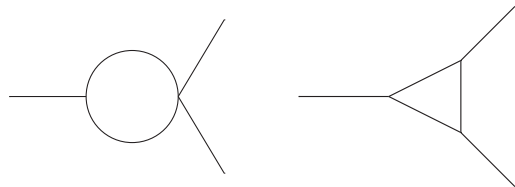
$$\begin{aligned}
 \frac{\vec{p}_1}{p_1} \text{---} \text{---} \frac{\leftarrow{p_2}}{p_2} &= -i(m^2 + p_1 \cdot p_2) + i \frac{\hbar}{16\pi^2} \frac{(24\lambda\phi)^2}{2} \left[-\ln(m^2) - p_1 \cdot p_2 \frac{1}{6m^2} + p_1^2 p_2^2 \frac{1}{60(m^2)^2} + \dots \right] \\
 &\quad + i \frac{\hbar}{16\pi^2} 12\lambda m^2 [1 - \ln m^2]
 \end{aligned}
 \tag{D13}$$

from which we extract

$$Z_2 = 1 + \frac{\hbar}{16\pi^2} 4\lambda \frac{12\lambda\phi^2}{m^2} = 1 + \frac{\hbar}{4\pi^2} \lambda, \tag{D14}$$

$$Z_4 = \frac{\hbar}{4\pi^2} \frac{\lambda}{10} \frac{12\lambda\phi^2}{(m^2)^2} = \frac{\hbar}{480\pi^2 \phi^2}. \tag{D15}$$

For the three-point function there are two types of one-loop diagrams,


(D16)

and for the four-point function there are three types of one-loop diagrams,

(D17)

Following the same procedure as for the two-point function, we find the three- and four-point function amplitudes

$$\begin{aligned}
 \begin{array}{c} \rightarrow \\ p_1 \end{array} \begin{array}{c} \swarrow \\ p_2 \end{array} \begin{array}{c} \nwarrow \\ p_3 \end{array} &= -i \left(24\lambda\phi + \frac{\hbar}{4\pi^2} 144\lambda^2\phi \left(1 + \frac{3}{2} \ln \frac{12\lambda\phi^2}{\mu^2} \right) \right) \\
 &\quad - i \left(\sum_{i>j} p_i^2 p_j^2 \right) \frac{\hbar}{240\pi^2\phi^3} - i \left(\sum_{\substack{i>j \\ k \neq i,j}} p_i \cdot p_j p_k^2 \right) \frac{\hbar}{720\pi^2\phi^3}
 \end{aligned} \tag{D18}$$

$$\begin{aligned}
 \begin{array}{c} \rightarrow \\ p_1 \end{array} \begin{array}{c} \swarrow \\ p_2 \end{array} \begin{array}{c} \nwarrow \\ p_3 \end{array} \begin{array}{c} \rightarrow \\ p_4 \end{array} &= -i \left(24\lambda + \frac{\hbar}{4\pi^2} (24\lambda)^2 \left(1 + \frac{3}{8} \ln \frac{12\lambda\phi^2}{\mu^2} \right) \right) + i \left(\sum_{i>j} p_i^2 p_j^2 \right) \frac{\hbar}{80\pi^2\phi^4} \\
 &\quad + i \left(\sum_{\substack{i>j \\ k \neq i,j}} p_i \cdot p_j p_k^2 \right) \frac{\hbar}{240\pi^2\phi^4} + i \left(\sum_{\substack{i>j, k>l \\ \text{all different}}} p_i \cdot p_j p_k \cdot p_l \right) \frac{\hbar}{360\pi^2\phi^4}
 \end{aligned} \tag{D19}$$

From these results we extract $\rho = \frac{\hbar}{720\pi^2\phi^3}$ and $\Omega = \frac{\hbar}{360\pi^2\phi^4}$.

To summarize our results, we found that, up to four derivatives, the effective action for ϕ^4 theory is given by

$$\begin{aligned}
 S_{\text{eff}}[\phi] &= \int d^4x \left[- \left(\lambda\phi^4 + \frac{9\hbar}{4\pi^2} \lambda^2\phi^4 \left(\ln \frac{12\lambda\phi^2}{\mu^2} - \frac{3}{2} \right) \right) + \frac{1}{2} (\partial_\mu\phi)^2 \left(1 + \frac{\hbar}{4\pi^2} \lambda \right) + \frac{1}{2} (\Box\phi)^2 \frac{\hbar}{480\pi^2} \frac{1}{\phi^2} \right. \\
 &\quad \left. - \frac{1}{2} (\partial_\mu\phi)^2 \Box\phi \frac{\hbar}{720\pi^2} \frac{1}{\phi^3} + \frac{1}{8} (\partial_\mu\phi)^2 (\partial_\nu\phi)^2 \frac{\hbar}{360\pi^2\phi^4} + \mathcal{O}(\partial^6) \right].
 \end{aligned} \tag{D20}$$

This is written in Minkowski space. Going to Euclidean space, we send $t \rightarrow -i\tau$, which changes $(\partial_\mu\phi)^2 \rightarrow -(\partial_\mu\phi)^2$ and $\Box\phi \rightarrow -\Box\phi$. Pulling out an extra minus sign (since $iS_{\text{eff}} = -S_{\text{eff}}^E$), we find

$$\begin{aligned}
 S_{\text{eff}}^E[\phi] &= \int d^4x \left[\lambda\phi^4 + \frac{9\hbar}{4\pi^2} \lambda^2\phi^4 \left(\ln \frac{12\lambda\phi^2}{\mu^2} - \frac{3}{2} \right) + \frac{1}{2} (\partial_\mu\phi)^2 \left(1 + \frac{\hbar}{4\pi^2} \lambda \right) - \frac{1}{2} (\Box\phi)^2 \frac{\hbar}{480\pi^2} \frac{1}{\phi^2} \right. \\
 &\quad \left. + \frac{1}{2} (\partial_\mu\phi)^2 \Box\phi \frac{\hbar}{720\pi^2} \frac{1}{\phi^3} - \frac{1}{8} (\partial_\mu\phi)^2 (\partial_\nu\phi)^2 \frac{\hbar}{360\pi^2\phi^4} + \mathcal{O}(\partial^6) \right]
 \end{aligned} \tag{D21}$$

where $d^4x = d\tau d^3\vec{x}$ and $\mu = 0$ corresponds to $x^0 = \tau$.

Our final result agrees with [67], where the four-derivative terms were computed using a different method.

APPENDIX E: NLO FUNCTIONAL DETERMINANTS

In quantum field theory, the calculation of decay rates at NLO amounts to evaluating the ratio of functional determinants:

$$\frac{\Gamma^{\text{NLO}}}{V} = \frac{1}{VT} \frac{e^{-S_E[\vec{\phi}]}}{e^{-S_E[\phi_{\text{FV}}]}} \left| \frac{\text{Det}(-\partial^2 + V''[\vec{\phi}(x)])}{\text{Det}(-\partial^2 + V''[\phi_{\text{FV}}])} \right|^{-1/2}. \tag{E1}$$

Here, $\bar{\phi}$ is the bounce solution to the Euclidean equations of motion, and ϕ_{FV} is the field value at the false vacuum which we can always choose to be zero. Generally, $\bar{\phi}(x)$ is spherically symmetric, only depending on the Euclidean length r . So we define

$$W(r) = V''[\bar{\phi}(r)]. \quad (\text{E2})$$

One can evaluate the functional determinant in terms of Feynman diagrams

$$\Delta S = \frac{1}{2} \log \text{Det}(-\partial^2 + W(r)) - \frac{1}{2} \log \text{Det}(-\partial^2) \quad (\text{E3})$$

$$= \frac{1}{2} \text{Tr} \log(-\partial^2 + W(r)) - \frac{1}{2} \text{Tr} \log(-\partial^2) \quad (\text{E4})$$

$$= -\text{Tr} \sum_{n=1}^{\infty} \frac{(-1)^n}{2n} [(-\partial^2)^{-1} W(r)]^n \quad (\text{E5})$$

$$= \tilde{W}(0) \otimes \text{circle} + \tilde{W}(q) \otimes \text{circle} \otimes \tilde{W}(-q) + \tilde{W}(p+q) \otimes \text{circle} \otimes \tilde{W}(-p) + \dots \quad (\text{E6})$$

$$= \frac{1}{2} \tilde{W}(0) \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2} - \frac{1}{4} \int \frac{d^d q}{(2\pi)^d} \frac{d^d k}{(2\pi)^d} \frac{\tilde{W}(q) \tilde{W}(-q)}{k^2 (k+q)^2} + \dots \quad (\text{E7})$$

We see that the first two terms are UV divergent, and these divergences can easily be removed using $\overline{\text{MS}}$ counterterms. All the other terms will be UV finite, but they are unfortunately very complicated to calculate. Note that we are not expanding in any small parameter, so, in general, all terms will be equally important. Hence, they are not only hard to calculate, but we would also have to calculate infinitely many of them.

There is an alternative way of calculating the functional determinants using the Gelfand-Yaglom theorem [15,57] which makes it possible to calculate ΔS to all orders in W . Since $W(r)$ only depends on the Euclidean distance r , we can decompose $[-\partial^2 + W(r)]$ into partial waves. We start by writing $\partial^2 = \frac{d^2}{dr^2} + \frac{3}{r} \frac{d}{dr} - \frac{L^2}{r^2} \equiv \nabla_l^2$, where L^2 is the four-dimensional angular momentum operator with eigenvalue $l(l+2)$ and degeneracy $(l+1)^2$ for $l = 0, 1, 2, \dots$. We can then write

$$\begin{aligned} \Delta S &\equiv \frac{1}{2} \ln \frac{\text{Det}[-\partial^2 + W(r)]}{\text{Det}[-\partial^2]} \\ &= \frac{1}{2} \sum_l (l+1)^2 \ln \frac{\text{Det}[-\nabla_l^2 + W(r)]}{\text{Det}[-\nabla_l^2]}. \end{aligned} \quad (\text{E8})$$

We can solve for the ratio of determinants for each l by solving for the two functions $u'_W(r)$ and $u'_0(r)$, which are

radial eigenfunctions of $-\partial^2 + W$ and $-\partial^2$, respectively, regular at $r = 0$, with a given l and zero eigenvalues. Then

$$\rho_l \equiv \frac{\text{Det}[-\nabla_l^2 + W(r)]}{\text{Det}[-\nabla_l^2]} = \lim_{r \rightarrow \infty} \rho_l(r) = \lim_{r \rightarrow \infty} \frac{u'_W(r)}{u'_0(r)}. \quad (\text{E9})$$

One can solve for $u'_0(r)$ and express the differential equation for $u'_W(r)$ in terms of $\rho_l(r)$,

$$\rho_l''(r) + \frac{2l+3}{r} \rho_l'(r) = W(r) \rho_l(r). \quad (\text{E10})$$

In summary, to calculate ΔS we need to find $\rho_l(r)$ from Eq. (E10) for each l and then sum the asymptotic values

$$\Delta S = \frac{1}{2} \sum_{l=0}^{\infty} (l+1)^2 \ln \rho_l, \quad (\text{E11})$$

where ρ_l is related to $\rho_l(r)$ by Eq. (E9).

There are two complications with directly implementing this approach. First, for most cases of interest, ρ_l can be either negative or zero for $l = 0, 1$, so that $\ln \rho_l$ is infinite or complex. The negative eigenvalue is of course expected, as the imaginary part is supposed to give the decay rate. The zero eigenvalues are also expected, as they correspond to exact or approximate symmetries such as translation or

scale invariance. See [15,20,57] for a discussion on how to evaluate the zero and negative eigenvalues.

Second, the sum over l is divergent, as we already knew it had to be from Eq. (E7). The UV divergence in Eq. (E7) came from the $\mathcal{O}(W)$ and $\mathcal{O}(W^2)$ terms which we can easily calculate analytically. So let us define $\Delta S^{[2]} \equiv [\Delta S]_{\mathcal{O}(W^2)}$, i.e. formally truncated to second order in W . We then add and subtract $\Delta S^{[2]}$, as well as subtracting off the infinities using $\overline{\text{MS}}$ counterterms,

$$\Delta S \rightarrow [\Delta S - \Delta S^{[2]}] + [\Delta S^{[2]} - \delta S_{\text{ct}}]. \quad (\text{E12})$$

The terms in the second set of brackets can be calculated analytically using Eq. (E7), and the terms in the first set of brackets will be calculated numerically.

To find ρ_j truncated to second order in W , we write $\rho_l(r) = 1 + \rho_l^{(1)}(r) + \rho_l^{(2)}(r) + \dots$, where $\rho_l^{(1)}(r) = \mathcal{O}(W)$, $\rho_l^{(2)}(r) = \mathcal{O}(W^2)$, etc. Solving Eq. (E10) order by order in W , we find the set of equations

$$\rho_l^{(1)''}(r) + \frac{2l+3}{r} \rho_l^{(1)'}(r) = W(r), \quad (\text{E13})$$

$$\rho_l^{(2)''}(r) + \frac{2l+3}{r} \rho_l^{(2)'}(r) = W(r) \rho_l^{(1)}(r). \quad (\text{E14})$$

When we have obtained these solutions, we can calculate

$$\Delta S^{[2]} = \frac{1}{2} \sum_{l=0}^{\infty} (l+1)^2 [\ln \rho_l]_{\mathcal{O}(W^2)} \quad (\text{E15})$$

$$= \frac{1}{2} \sum_{l=0}^{\infty} (l+1)^2 [\ln(1 + \rho_l^{(1)} + \rho_l^{(2)} + \dots)]_{\mathcal{O}(W^2)} \quad (\text{E16})$$

$$= \frac{1}{2} \sum_{l=0}^{\infty} (l+1)^2 \left[\rho_l^{(1)} - \frac{1}{2} (\rho_l^{(1)})^2 + \rho_l^{(2)} \right]. \quad (\text{E17})$$

Hence, we find that the numerical bracket is

$$[\Delta S - \Delta S^{[2]}] = \frac{1}{2} \sum_{l=0}^{\infty} (l+1)^2 \left[\ln \rho_l - \rho_l^{(1)} + \frac{1}{2} (\rho_l^{(1)})^2 - \rho_l^{(2)} \right] \quad (\text{E18})$$

which will be finite as $l \rightarrow \infty$. In practice, this sum rapidly converges, and one only has to sum a finite number of terms.

We can simplify $[\Delta S^{[2]} - \delta S_{\text{ct}}]$ by noting that the $\mathcal{O}(W)$ term in Eq. (E7) has a scaleless momentum space integral, which is zero in dimensional regularization using $\overline{\text{MS}}$. The $\mathcal{O}(W^2)$ integral can be simplified by first doing an integral over k ,

$$B_0(q^2) = \mu^{4-d} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2(k+q)^2} = \frac{1}{(4\pi)^2} \left[\frac{1}{\epsilon} + 2 + \ln \frac{\mu^2}{q^2} \right]. \quad (\text{E19})$$

Removing the infinity using the counterterm, we find

$$[\Delta S^{[2]} - \delta S_{\text{ct}}] = -\frac{1}{4} \int \frac{d^4 q}{(2\pi)^4} [\tilde{W}(q)]^2 B_0(q^2). \quad (\text{E20})$$

In the case of $V = -\frac{|\lambda|}{4!} \phi^4$,

$$\begin{aligned} W(r) &= -\frac{|\lambda|}{2} \bar{\phi}(r)^2 \\ &= -\frac{24R^2}{(r^2 + R^2)^2} \Rightarrow \tilde{W}(p) \\ &= 48\pi^2 R^2 K_0(|p|R) \end{aligned} \quad (\text{E21})$$

where K_0 is a BesselK function. Using Eq. (E20), we then find

$$[\Delta S^{[2]} - \delta S_{\text{ct}}] = -3L - \frac{5}{2} \quad (\text{E22})$$

where $L = \ln \frac{R\mu e^{\gamma_E}}{2}$.

-
- [1] J. Callan, G. Curtis, and S. R. Coleman, The fate of the false vacuum. 2. First quantum corrections, *Phys. Rev. D* **16**, 1762 (1977).
[2] I. Kobzarev, L. B. Okun, and M. B. Voloshin, Bubbles in metastable vacuum, *Yad. Fiz.* **20**, 1229 (1974) [*Sov. J. Nucl. Phys.* **20**, 644 (1975)].
[3] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, 5th ed. (World Scientific Publishing, Singapore, 2009).
[4] H. J. W. Müller-Kirsten, *Introduction to Quantum Mechanics* (World Scientific, Singapore, 2012).

- [5] J. Zinn-Justin, *Quantum field theory and critical phenomena*, Int. ser. monogr. phys. **113**, 1 (2002).
[6] M. Marino, *Instantons and Large N: An Introduction to Non-Perturbative Methods in Quantum Field Theory* (Cambridge University Press, Cambridge, England, 2015).
[7] E. J. Weinberg, *Classical Solutions in Quantum Field Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2012).
[8] A. Andreassen, D. Farhi, W. Frost, and M. D. Schwartz, A Direct Approach to Quantum Tunneling, *Phys. Rev. Lett.* **117**, 231601 (2016).

- [9] T. Banks, C. M. Bender, and T. T. Wu, Coupled anharmonic oscillators. 1. Equal mass case, *Phys. Rev. D* **8**, 3346 (1973).
- [10] S. R. Coleman and E. J. Weinberg, Radiative corrections as the origin of spontaneous symmetry breaking, *Phys. Rev. D* **7**, 1888 (1973).
- [11] S. R. Coleman, Fate of the false vacuum, *Subnuclear series* **15**, 805 (1979).
- [12] P. H. Frampton, Consequences of vacuum instability in quantum field theory, *Phys. Rev. D* **15**, 2922 (1977).
- [13] M. Sher, Electroweak Higgs potentials and vacuum stability, *Phys. Rep.* **179**, 273 (1989).
- [14] J. Espinosa and M. Quiros, Improved metastability bounds on the standard model Higgs mass, *Phys. Lett. B* **353**, 257 (1995).
- [15] G. Isidori, G. Ridolfi, and A. Strumia, On the metastability of the standard model vacuum, *Nucl. Phys. B* **609**, 387 (2001).
- [16] J. Espinosa, G. Giudice, and A. Riotto, Cosmological implications of the Higgs mass measurement, *J. Cosmol. Astropart. Phys.* **05** (2008) 002.
- [17] J. Ellis, J. Espinosa, G. Giudice, A. Hoecker, and A. Riotto, The probable fate of the standard model, *Phys. Lett. B* **679**, 369 (2009).
- [18] L. Di Luzio, G. Isidori, and G. Ridolfi, Stability of the electroweak ground state in the Standard Model and its extensions, *Phys. Lett. B* **753**, 150 (2016).
- [19] V. Branchina, E. Messina, and M. Sher, Lifetime of the electroweak vacuum and sensitivity to Planck scale physics, *Phys. Rev. D* **91**, 013003 (2015).
- [20] V. Branchina and E. Messina, Stability, Higgs Boson Mass and New Physics, *Phys. Rev. Lett.* **111**, 241801 (2013).
- [21] V. Branchina, Stability of the EW vacuum, Higgs boson, and new physics, *arXiv:1405.7864*.
- [22] V. Branchina, E. Messina, and A. Platania, Top mass determination, Higgs inflation, and vacuum stability, *J. High Energy Phys.* **09** (2014) 182.
- [23] G. Gamow, Quantum theory of the atomic nucleus, *Z. Phys.* **51**, 204 (1928).
- [24] D. Griffiths, *Introduction to Quantum Mechanics*, Pearson International Edition (Pearson Prentice Hall, London, England, 2005).
- [25] W. van Dijk and Y. Nogami, Novel Expression for the Wave Function of a Decaying Quantum System, *Phys. Rev. Lett.* **83**, 2867 (1999).
- [26] R. de la Madrid and M. Gadella, A pedestrian introduction to Gamow vectors, *Am. J. Phys.* **70**, 626 (2002).
- [27] S. Flügge, *Nuclear Reactions II: Theory/Kernreaktionen II: Theorie*, Handbuch der Physik Encyclopedia of Physics (Springer, Berlin, Heidelberg, 1959).
- [28] B. R. Holstein and A. R. Swift, Barrier penetration via path integrals, *Am. J. Phys.* **50**, 833 (1982).
- [29] B. R. Holstein, Bound states, virtual states, and nonexponential decay via path integrals, *Am. J. Phys.* **51**, 897 (1983).
- [30] M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964).
- [31] R. G. Newton, *Scattering Theory of Particles and Waves* (Springer, New York, 1982).
- [32] U. D. Jentschura, A. Surzhykov, and J. Zinn-Justin, Multi-instantons and exact results III: Unification of even and odd anharmonic oscillators, *Ann. Phys. (Amsterdam)* **325**, 1135 (2010).
- [33] U. D. Jentschura and J. Zinn-Justin, Multi-instantons and exact results. IV: Path integral formalism, *Ann. Phys. (Amsterdam)* **326**, 2186 (2011).
- [34] C. M. Bender and T. T. Wu, Anharmonic oscillator. 2: A Study of perturbation theory in large order, *Phys. Rev. D* **7**, 1620 (1973).
- [35] J. C. Collins and D. E. Soper, Large order expansion in perturbation theory, *Ann. Phys. (N.Y.)* **112**, 209 (1978).
- [36] E. Witten, Analytic continuation of Chern-Simons theory, *AMS/IP Stud. Adv. Math.* **50**, 347 (2011).
- [37] A. Behtash, G. V. Dunne, T. Schaefer, T. Sulejmanpasic, and M. Unsal, Complexified Path Integrals, Exact Saddles and Supersymmetry, *Phys. Rev. Lett.* **116**, 011601 (2016).
- [38] C. L. Hammer, J. E. Shrauner, and B. DeFacio, Alternate derivation of vacuum tunneling, *Phys. Rev. D* **19**, 667 (1979).
- [39] S. R. Coleman, Quantum tunneling and negative eigenvalues, *Nucl. Phys. B* **298**, 178 (1988).
- [40] S. R. Coleman, The fate of the false vacuum. 1. Semi-classical theory, *Phys. Rev. D* **15**, 2929 (1977).
- [41] J.-L. Gervais and B. Sakita, Extended particles in quantum field theories, *Phys. Rev. D* **11**, 2943 (1975).
- [42] J.-L. Gervais, A. Jevicki, and B. Sakita, Perturbation expansion around extended particle states in quantum field theory. 1., *Phys. Rev. D* **12**, 1038 (1975).
- [43] C. G. Callan, Jr. and D. J. Gross, Quantum perturbation theory of solitons, *Nucl. Phys. B* **93**, 29 (1975).
- [44] A. Jevicki, Treatment of zero frequency modes in perturbation expansion about classical field configurations, *Nucl. Phys. B* **117**, 365 (1976).
- [45] J. Zinn-Justin, *Path Integrals in Quantum Mechanics* (Oxford University Press, New York, 2005).
- [46] K.-M. Lee and E. J. Weinberg, Tunneling without barriers, *Nucl. Phys. B* **267**, 181 (1986).
- [47] A. R. Brown and A. Dahlen, The case of the disappearing instanton, *Phys. Rev. D* **84**, 105004 (2011).
- [48] S. R. Coleman, V. Glaser, and A. Martin, Action minima among solutions to a class of Euclidean scalar field equations, *Commun. Math. Phys.* **58**, 211 (1978).
- [49] A. Ferraz de Camargo, R. C. Shellard, and G. C. Marques, Vacuum decay in a soluble model, *Phys. Rev. D* **29**, 1147 (1984).
- [50] K. Dutta, C. Hector, P. M. Vaudrevange, and A. Westphal, More exact tunneling solutions in scalar field theory, *Phys. Lett. B* **708**, 309 (2012).
- [51] A. Aravind, B. S. DiNunno, D. Lorshbough, and S. Paban, Analyzing multifield tunneling with exact bounce solutions, *Phys. Rev. D* **91**, 025026 (2015).
- [52] D. A. Samuel and W. A. Hiscock, ‘Thin wall’ approximations to vacuum decay rates, *Phys. Lett. B* **261**, 251 (1991).
- [53] T. C. Shen, Bubbles without cores, *Phys. Rev. D* **37**, 3537 (1988).
- [54] E. J. Weinberg, Vacuum decay in theories with symmetry breaking by radiative corrections, *Phys. Rev. D* **47**, 4614 (1993).

- [55] E. J. Weinberg and A.-q. Wu, Understanding complex perturbative effective potentials, *Phys. Rev. D* **36**, 2474 (1987).
- [56] A. D. Plascencia and C. Tamarit, Convexity, gauge-dependence and tunneling rates, *J. High Energy Phys.* **10** (2016) 099.
- [57] G. V. Dunne and H. Min, Beyond the thin-wall approximation: Precise numerical computation of prefactors in false vacuum decay, *Phys. Rev. D* **72**, 125004 (2005).
- [58] C. M. Bender, F. Cooper, B. Freedman, and R. W. Haymaker, Tunneling and the low momentum expansion of the effective action, *Nucl. Phys.* **B256**, 653 (1985).
- [59] A. Andreassen, W. Frost, and M. D. Schwartz, Consistent use of effective potentials, *Phys. Rev. D* **91**, 016009 (2015).
- [60] D. Metaxas and E. J. Weinberg, Gauge independence of the bubble nucleation rate in theories with radiative symmetry breaking, *Phys. Rev. D* **53**, 836 (1996).
- [61] D. Buttazzo, G. Degrandi, P. P. Giardino, G. F. Giudice, F. Sala *et al.*, Investigating the near-criticality of the Higgs boson, *J. High Energy Phys.* **12** (2013) 089.
- [62] N. Nielsen, On the gauge dependence of spontaneous symmetry breaking in gauge theories, *Nucl. Phys.* **B101**, 173 (1975).
- [63] R. Fukuda and T. Kugo, Gauge invariance in the effective action and potential, *Phys. Rev. D* **13**, 3469 (1976).
- [64] A. Andreassen, W. Frost, and M. D. Schwartz, Consistent Use of the Standard Model Effective Potential, *Phys. Rev. Lett.* **113**, 241801 (2014).
- [65] M. Razavy, *Quantum Theory of Tunneling*, 2nd ed. (World Scientific, Singapore, 2013).
- [66] J.-L. Gervais, A. Jevicki, and B. Sakita, Collective coordinate method for quantization of extended systems, *Phys. Rep.* **23**, 281 (1976).
- [67] C. Fraser, Calculation of higher derivative terms in the one loop effective Lagrangian, *Z. Phys. C* **28**, 101 (1985).