

WKB theory of epidemic fade-out in stochastic populations

Baruch Meerson¹ and Pavel V. Sasorov²

¹*Racah Institute of Physics, Hebrew University of Jerusalem, Jerusalem 91904, Israel*

²*Institute of Theoretical and Experimental Physics, Moscow 117218, Russia*

(Received 25 June 2009; revised manuscript received 5 September 2009; published 27 October 2009)

Stochastic effects may cause “fade-out” of an infectious disease in a population immediately after an epidemic outbreak. We evaluate the epidemic fade-out probability by a WKB method and find that the most probable path to extinction of the disease comes from an instantonlike orbit in the phase space of an underlying Hamiltonian flow.

DOI: [10.1103/PhysRevE.80.041130](https://doi.org/10.1103/PhysRevE.80.041130)

PACS number(s): 02.50.Ga, 87.23.Cc, 87.10.Mn

An infectious disease can disappear from a population immediately after a major infection outbreak [1,2]. This phenomenon, called “epidemic fade-out,” occurs if the epidemic dynamics is oscillatory and the number of infected individuals at the end of the first outbreak of the disease is relatively low so that fluctuations in the disease transmission can “switch off” the disease. Epidemic fade-out has been addressed by epidemiologists via stochastic simulations. One exception is Ref. [2] which arrived at important analytical results which we briefly review and use below.

Epidemic fade-out is an example of a large fluctuation in a multivariate stochastic system far from equilibrium. There is no general theory of fluctuations in such systems and finding the probability of a large fluctuation is hard. Here we develop a theoretical framework for analysis of epidemic fade-out on the example of stochastic *SI* model: a Markov process involving susceptible and infected subpopulations [1–3]. We formulate the problem in a master equation setting. In contrast to *endemic* fade-out, which can be studied assuming quasistationarity [4], the epidemic fade-out occurs on a *fast* time scale (determined by the deterministic rate equations of the *SIR* model), so no ready-to-use methods of solution are available. To overcome this difficulty we derive a *stationary* equation for the mean residence time of the population in a certain state. Then we develop a WKB theory using the population size as a large parameter. In the WKB framework the problem reduces to that of finding a special zero-energy phase orbit of the underlying Hamiltonian: the orbit which provides a global minimum to the action functional for boundary conditions corresponding to epidemic fade-out. This orbit, which encodes the most probable path of the population toward the disease extinction, turns out to be instantonlike [5]. We observe that the epidemic fade-out instanton exists only in the regime when the epidemic dynamics, as described by the deterministic rate equations, is oscillatory. Of special interest is the regime where the number of infected exhibits *large* oscillations prior to reaching the endemic state. By using a matched asymptotic expansion, we analytically calculate the action along the instanton which determines the epidemic fade-out probability.

The probability $P_{n,m}(t)$ to observe, at time t , n susceptible and m infected individuals is governed by the master equation with transition rates from Table I. Solving the master equation analytically is hard. The often used van Kampen system-size expansion (vKSSE) [6] approximates the master equation by a Fokker-Planck equation. In the context of epi-

dem fade-out in the *SI* model the vKSSE was employed in Ref. [2]. The vKSSE is very useful for “typical” fluctuations [6], but it often fails for large fluctuations [7], such as those causing extinction.

Before dealing with the master equation, consider the deterministic *rate equations* for the *SI* model:

$$\dot{S} = \mu N - (\beta/N)SI - \mu S, \quad \dot{I} = (\beta/N)SI - \Gamma I. \quad (1)$$

For $\beta > \Gamma$ there is an attracting fixed point $\bar{S} = (\Gamma/\beta)N$, $\bar{I} = \mu(1/\Gamma - 1/\beta)N$ which describes an endemic infection level, and an unstable (saddle) point $\bar{S} = N$, $\bar{I} = 0$ which describes an infection-free population. At $\mu > 4(\beta - \Gamma)(\Gamma/\beta)^2$ the attracting fixed point is a stable node. We are mostly interested in the opposite inequality, when the attracting fixed point is a stable focus and the epidemic dynamics is oscillatory. Let a few infected be introduced into a susceptible population. For small μ the minimum number of infected at the end of the first outbreak of the disease is small; see the dashed line in Fig. 1. As a result, stochasticity, missed by the rate equations, can switch off the disease before the endemic level is reached. To describe this process we use the master equation

$$\begin{aligned} \dot{P}_{n,m} = & \sum_{n',m'} M_{n,m;n',m'} P_{n',m'}(t) = \mu [N(P_{n-1,m} - P_{n,m}) \\ & + (n+1)P_{n+1,m} - nP_{n,m}] + \Gamma [(m+1)P_{n,m+1} - mP_{n,m}] \\ & + (\beta/N)[(n+1)(m-1)P_{n+1,m-1} - nmP_{n,m}]. \end{aligned} \quad (2)$$

A natural initial condition is a product of Kronecker deltas: $P_{n,m}(t=0) = \delta_{n,N} \delta_{m,m_0}$. One boundary condition reflects the fact that $m=0$ is, for any n , an absorbing state. Being interested in epidemic fade-out, we exclude from consideration all stochastic trajectories that do not reach the extinction boundary $m=0$ immediately after the first outbreak and leave the region of small m . This is achieved by introducing an

TABLE I. Stochastic *SI* model.

| Event | Type of transition | Rate |
|------------------------|----------------------------------------|---------------|
| Infection | $S \rightarrow S-1, I \rightarrow I+1$ | $(\beta/N)SI$ |
| Renewal of susceptible | $S \rightarrow S+1$ | μN |
| Removal of infected | $I \rightarrow I-1$ | ΓI |
| Removal of susceptible | $S \rightarrow S-1$ | μS |

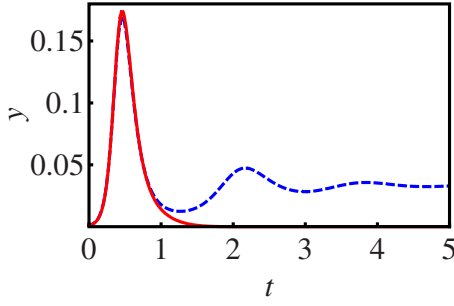


FIG. 1. (Color online) An epidemic outbreak in the SI model. Shown is the rescaled number of infected $y=I/N$ versus rescaled time μt . Dashed line: prediction from rate equations (5). Solid line: the epidemic fade-out instanton. The rescaled parameters $K=\beta/\mu=30$ and $\delta=1-\Gamma/\beta=0.5$.

artificial absorbing boundary [6] that will be specified shortly.

The disease can only disappear from the population via transition from a state $(n, 1)$ to the state $(n, 0)$. Consider the mean residence time $T_{n,m} = \int_0^\infty P_{n,m}(t) dt$ of the system in the state (n, m) , where $m > 0$. The accumulated extinction probability \mathcal{P}_n from the state $(n, 1)$ is $\mathcal{P}_n = \Gamma T_{n,1}$ and the total extinction probability is $\mathcal{P} = \sum_n \mathcal{P}_n$. Integrating Eq. (2) over t from 0 to ∞ and using the equality $P_{n,m}(t=\infty) = 0$ and the initial condition, we obtain a stationary equation for $T_{n,m>0}$:

$$\sum_{n',m'>0} M_{n,m;n',m'} T_{n',m'} + \delta_{n,n_0} \delta_{m,m_0} = 0. \quad (3)$$

We assume throughout this work that $N \gg 1$. Here the stochasticity is weak (but very important), and Eq. (3) can be approximately solved by the WKB ansatz $T_{n,m} = a(x, y) e^{-NS(x,y)}$ [8], where a and S are smooth functions of the continuous variables $x = n/N - 1$ and $y = m/N$.

In the leading WKB order one arrives at a stationary Hamilton-Jacobi equation $H(x, y, \partial_x S, \partial_y S) = 0$, where

$$H(x, y, p_x, p_y) = e^{p_x} - 1 + (1+x)(e^{-p_x} - 1) + K(1-\delta)y(e^{-p_y} - 1) + Ky(1+x)(e^{p_y - p_x} - 1), \quad (4)$$

and we have introduced rescaled parameters $\delta = 1 - \Gamma/\beta$ and $K = \beta/\mu$ and rescaled time by the rate constant μ [9]. The four-dimensional (4D) phase space, defined by Hamiltonian (4), yields an instructive visualization of the most probable path of the disease toward fade-out. As H is independent of time, $H(x, y, p_x, p_y) = E = \text{const}$. Furthermore, in view of stationarity of the Hamilton-Jacobi equation, we only need to deal with zero-energy orbits, $E = 0$. The simplest among them are fluctuationless orbits lying in the plane $p_x = p_y = 0$. These are described by the equations

$$\dot{x} = -x - Ky(1+x), \quad \dot{y} = -K(1-\delta)y + Ky(1+x), \quad (5)$$

which coincide with (rescaled) rate equations (1). Disease fade-out demands a fluctuational orbit, for which the momenta p_x and p_y are nonzero. Before dealing with such orbits, consider the fixed points of the zero-energy Hamiltonian. There are exactly three such points, all of them 4D

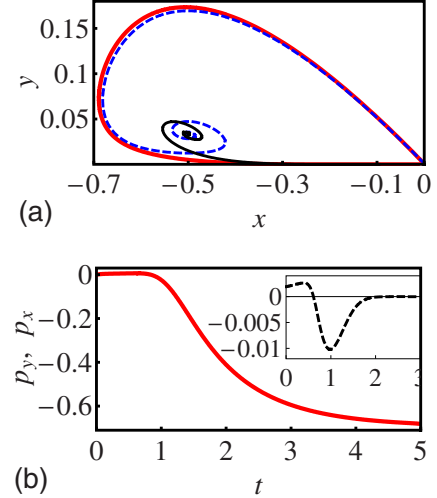


FIG. 2. (Color online) (a) An epidemic outbreak on the xy plane as predicted by rate equations (5) (dashed line) and the epidemic fade-out instanton (thick solid line). Also shown is the endemic fade-out instanton [4] (thin solid line). (b) p_y (inset: p_x) vs t for the epidemic fade-out instanton. The rescaled parameters are $K=30$ and $\delta=1/2$.

saddles [4]. Two of them, $B = [0, 0, 0, \ln(1-\delta)]$ and $C = [0, 0, 0, 0]$, describe infection-free steady states. Point C is fluctuationless: it corresponds to the saddle point of the rate equations. Point B is fluctuational, as its $p_y \neq 0$. Finally, the fluctuationless fixed point $A = [-\delta, (\delta/K)(1-\delta)^{-1}, 0, 0]$ corresponds to the endemic fixed point of the rate equations.

Let one or few infected be introduced into an infection-free population. In the leading WKB order this initial condition can correspond to different phase-space points whose projections on the xy plane are very close to the fluctuationless fixed point C . Each of these phase-space points generates an orbit which exits the fixed point C along the manifold spanned by its two unstable eigenvectors. For epidemic fade-out to occur, such an orbit must reach the extinction hyperplane $y=0$ before crossing, say, the hyperplane $y = -(x/K)(1-\delta)^{-1}$, $-\delta < x < 0$ (which is a 4D extension of the artificial absorbing boundary mentioned above). One can prove that, among all such orbits, the one providing the global minimum to the action $S(x, 0)$ (and therefore the global maximum to the fade-out probability \mathcal{P}_n) ends in the fluctuational fixed point B . As a result, $\max \mathcal{P}_n = \mathcal{P}_N$. Therefore, at $N \gg 1$, the epidemic fade-out problem reduces to that of finding a heteroclinic orbit going from C to B . We found numerically that such a heteroclinic orbit CB exists if and only if $K > K_c = (1/4\delta)(1-\delta)^{-2}$: when the endemic fixed point, predicted by the rate equations, is a focus. As K exceeds K_c , the heteroclinic orbit emerges via a global bifurcation. In fact, one finds multiple heteroclinic orbits at $K > K_c$. They can be classified by whether their xy projections exhibit a single loop, two loops, three loops, etc. A single-loop orbit, see Fig. 2, corresponds to disease fade-out immediately after the first outbreak. A two-loop orbit corresponds to a fade-out immediately after the second outbreak, etc. The connection between the rapid epidemic fade-out in a stochastic population and a zero-energy instantonlike orbit of an effective Hamiltonian is a central result of our work. We stress that previous

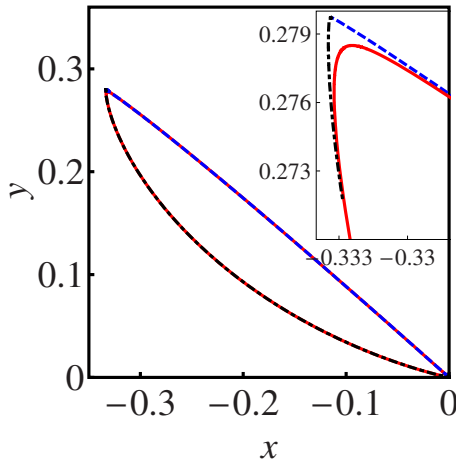


FIG. 3. (Color online) Same as in Fig. 2 but for $K=1.7875$ and $\delta=1/3$, so $K_c=1.6875$. The endemic fade-out instanton [4] is shown by the dash-dotted line. Inset: a blowup near the endemic fixed point A.

studies, which found instantonlike orbits in the context of population extinction, dealt with extinction from *long-lived metastable* states.

How does the epidemic fade-out instanton look like at different parameters? For $K\delta \gg 1$ the fraction of infected versus time, $y(t)$, first rapidly grows and becomes large and then falls down to a small value (see Fig. 1, solid line), closely following the prediction from the rate equations. Then $y(t)$ strongly deviates from the deterministic path and rapidly goes to zero. The x , p_x , and p_y dynamics for the same values of K and δ are depicted in Fig. 2. One can see that a rapid deviation from the deterministic path occurs around $x=-\delta$. Notably, $|p_x|$ remains much smaller than unity *everywhere*. $|p_y|$, however, is steadily growing and, at nonsmall δ , becomes $\mathcal{O}(1)$ as the instanton approaches the fluctuational fixed point B. As a result, the vKSSE is invalid for most of the small- y region where disease extinction occurs.

Near the bifurcation, $0 < K - K_c \ll K_c$, our numerics reveals an intimate relation between the epidemic fade-out instanton and two other zero-energy heteroclinic orbits. The first is the *deterministic* orbit which lies in the xy plane and goes from C to A. The second is the endemic fade-out instanton: a heteroclinic orbit which goes from A to B and describes stochastic endemic fade-out [4]. The xy projection of the epidemic fade-out instanton initially closely follows the deterministic orbit CA; see Fig. 3. The momenta p_x and p_y are very small here. They slowly build up and become important only when the xy projection of the instanton reaches a close vicinity of the endemic point A. Here the projection of the epidemic fade-out instanton departs from the deterministic orbit (see the inset of Fig. 3) and rapidly approaches the projection of the endemic fade-out instanton.

To evaluate $\mathcal{P} \sim \mathcal{P}_N$ in the leading WKB order, we need to calculate the accumulated action \mathcal{S}_0 along the instanton. In the rest of this paper we will focus on the important regime of $K\delta \gg 1$, when the fade-out probability can indeed be significant. It turns out that the presence of the small parameter $(K\delta)^{-1}$ enables one to find the instanton, and calculate \mathcal{S}_0 , analytically. An immediate simplification comes from the

fact that the fluctuations of the number of susceptibles are negligible everywhere, so we can Taylor expand Hamiltonian (4) in $p_x \ll 1$ and truncate the expansion at first order. Another simplification employs the strong inequality $y \ll \delta$ which holds in the whole region where the epidemic fade-out instanton significantly deviates from the deterministic orbit. A complete calculation of the instanton will be reported elsewhere. Here we will analytically calculate \mathcal{S}_0 . As can be verified *a posteriori*, the main contribution to \mathcal{S}_0 comes from a narrow region $|x + \delta| \ll \delta$, where the instanton rapidly departs from the deterministic orbit. Furthermore, $|p_y| \ll 1$ in this narrow region, so one can Taylor expand Eq. (4) in p_y and truncate the expansion at p_y^2 . Neglecting small terms, we can reduce Hamiltonian (4) to

$$H(x, y, p_x, p_y) \simeq \delta p_x + K y p_y [x + \delta + (1 - \delta) p_y]. \quad (6)$$

The reduced problem is integrable. There is no need in the full solution, however, if one only needs to evaluate \mathcal{S}_0 . The Hamilton's equation for \dot{x} yields $x(t) = \delta(t-1)$, where the arbitrary constant is fixed by choosing $x(t=0) = -\delta$. The Hamilton's equation for \dot{p}_y reads

$$\dot{p}_y = -K p_y [x + \delta + (1 - \delta) p_y]. \quad (7)$$

Plugging here $x = \delta(t-1)$, we obtain an exactly soluble equation for $p_y(t)$. The boundedness of $p_y(t)$ fixes the integration constant and we obtain

$$p_y = \frac{1}{K(1-\delta)} \frac{d}{dt} \ln \int_t^\infty e^{-K\delta u^2/2} du. \quad (8)$$

Now let us calculate \dot{S} along the instanton: $\dot{S} = p_x \dot{x} + p_y \dot{y} = H + K(1-\delta)y p_y^2 \equiv \mathcal{F}$, where we have used $H=0$ and denoted $\mathcal{F} \equiv K(1-\delta)y p_y^2$. Using the Hamilton's equations, we observe that $\mathcal{F}(t)$ obeys the equation $\dot{\mathcal{F}} = -K(x+\delta)\mathcal{F} = -K\delta t \mathcal{F}$. Integration yields

$$\mathcal{F}(t) \equiv K(1-\delta)y(t)p_y^2(t) = C \exp(-K\delta t^2/2), \quad (9)$$

where $C = \text{const}$. Therefore, $\dot{S} = C \exp(-K\delta t^2/2)$ and

$$\mathcal{S}_0 = \int_{-\infty}^{\infty} \dot{S} dt = C \sqrt{\frac{2\pi}{K\delta}}. \quad (10)$$

What is left is to find C . Importantly, the deterministic solution still holds in the region of $-x - \delta \ll \delta$ (or $-t \ll 1$). For $K\delta \gg 1$ the deterministic solution was found by van Herwaarden [2]. In the region of $-x - \delta \ll \delta$, Eqs. (3.25 a-d) of van Herwaarden can be simplified and rewritten, in our notation, as

$$y(t) = y_m \exp(K\delta t^2/2), \quad (11)$$

$$y_m = y_m(K, \delta) = \frac{(\delta + x_m)x_m}{1 + x_m} \left(\frac{-x_m}{\delta} \right)^{K\delta} \times \exp[K(x_m + \delta) - (1 + x_m^{-1})\mathcal{Q}_1(x_m)], \quad (12)$$

where $x_m = x_m(\delta)$ is the negative root of the equation $x_m = (1-\delta)\ln(1+x_m)$, and $\mathcal{Q}_1(x_m)$ is given by

$$Q_1(x_m) = \int_0^{x_m} \left[\frac{s(s+\delta)}{(1+s)^2[s-(1-\delta)\ln(1+s)]} - \frac{x_m}{(1+x_m)(s-x_m)} \right] ds. \quad (13)$$

(For $\delta \rightarrow 0$ one obtains $x_m(\delta) \simeq -2\delta$ and $Q_1 \simeq -4\delta$.) In the region of $(K\delta)^{-1/2} \ll -x - \delta \ll \delta$ Eq. (8) becomes

$$p_y(t) = -(1-\delta)^{-1} [\delta/(2\pi K)]^{1/2} \exp(-K\delta t^2/2). \quad (14)$$

Using Eqs. (9), (11), and (14) in their joint validity region $(K\delta)^{-1/2} \ll -x - \delta \ll \delta$, we obtain $C = y_m \delta / [2\pi(1-\delta)]$. Putting everything together, we obtain the epidemic fade-out probability $\mathcal{P} \sim \exp(-N\mathcal{S}_0)$, where \mathcal{S}_0 is given by Eq. (10) and y_m is given by Eqs. (12) and (13). Note that \mathcal{S}_0 is exponentially small in $K\delta \gg 1$, so the WKB result holds only for very large N : $N\mathcal{S}_0 \gg 1$. In Fig. 4 our results for \mathcal{S}_0 are compared with those obtained by a numerical integration of the Hamilton's equations. For large $K\delta$ the agreement is very good. For $N\mathcal{S}_0 \lesssim 1$ the epidemic fade-out probability is large.

That truncation of H at p_y^2 yields, at $K\delta \gg 1$, an accurate leading-order result for \mathcal{S}_0 justifies the validity of the vKSSE for calculating \mathcal{S}_0 . Indeed, our leading-order result for \mathcal{P} coincides with that obtained, by a different method, by van Herwaarden [2] who used the vKSSE. We reiterate, however, that the vKSSE is invalid in most of the small- y region, while the full WKB Hamiltonian (4) holds there. Only at $\delta \ll 1$, when $|p_y| \ll 1$ on the whole instanton, the vKSSE de-

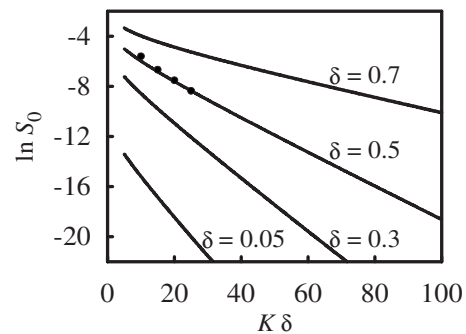


FIG. 4. Natural logarithm of the action \mathcal{S}_0 along the instanton versus $K\delta$ at different δ as predicted by our asymptotic theory at $K\delta \gg 1$ (lines) and found by a numerical integration of the full Hamilton's equations (circles) for $\delta=0.5$.

scribes the instanton correctly. In this case one obtains $\mathcal{S}_0 = (2\delta^5 / \pi e^4 K)^{1/2} (e/2)^{-K\delta}$.

In summary, we have shown that rapid epidemic fade-out in stochastic populations is amenable to an accurate analysis via a WKB theory. We calculated the epidemic fade-out probability and established an unexpected connection between the rapid epidemic fade-out and an instantonlike orbit of an underlying Hamiltonian. The epidemic fade-out instanton should be observable in stochastic simulations of, and actual data on, epidemics in small communities.

This work was supported by the Israel Science Foundation (Grant No. 408/08).

[1] M. S. Bartlett, *Stochastic Population Models in Ecology and Epidemiology* (Wiley, New York, 1961); N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications* (Griffin, London, 1975); R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).
 [2] O. A. van Herwaarden, *J. Math. Biol.* **35**, 793 (1997).
 [3] The closely related *SIR* model (where *R* stands for recovered) can be reduced to the *SI* model; see, e.g., Ref. [4].
 [4] A. Kamenev and B. Meerson, *Phys. Rev. E* **77**, 061107 (2008).
 [5] R. Rajaraman, *Solitons and Instantons* (North Holland, Amsterdam, 1987).
 [6] N. G. van Kampen, *Stochastic Processes in Physics and*

Chemistry (North-Holland, Amsterdam, 2001); C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2004).
 [7] B. Gaveau, M. Moreau, and J. Toth, *Lett. Math. Phys.* **37**, 285 (1996).
 [8] R. Kubo, K. Matsuo, and K. Kitahara, *J. Stat. Phys.* **9**, 51 (1973); M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, *J. Chem. Phys.* **100**, 5735 (1994).
 [9] Hamiltonian (4) is equivalent, up to a canonical transformation, to the Hamiltonian obtained in Ref. [4] where (a different variant of) WKB approximation was used for a study of epidemic fade-out.