

**Outbreak Size Distribution in Stochastic Epidemic Models**Jason Hindes,<sup>1</sup> Michael Assaf<sup>2,3</sup>, and Ira B. Schwartz<sup>1</sup><sup>1</sup>*U.S. Naval Research Laboratory, Washington, D.C. 20375, USA*<sup>2</sup>*Racah Institute of Physics, Hebrew University of Jerusalem, Jerusalem 91904, Israel*<sup>3</sup>*Institute for Physics and Astronomy, University of Potsdam, Potsdam 14476, Germany* (Received 27 July 2021; revised 10 November 2021; accepted 26 January 2022; published 15 February 2022)

Motivated by recent epidemic outbreaks, including those of COVID-19, we solve the canonical problem of calculating the dynamics and likelihood of extensive outbreaks in a population within a large class of stochastic epidemic models with demographic noise, including the susceptible-infected-recovered (SIR) model and its general extensions. In the limit of large populations, we compute the probability distribution for all extensive outbreaks, including those that entail unusually large or small (extreme) proportions of the population infected. Our approach reveals that, unlike other well-known examples of rare events occurring in discrete-state stochastic systems, the statistics of extreme outbreaks emanate from a full continuum of Hamiltonian paths, each satisfying unique boundary conditions with a conserved probability flux.

DOI: [10.1103/PhysRevLett.128.078301](https://doi.org/10.1103/PhysRevLett.128.078301)

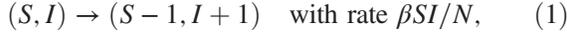
*Introduction.*—Epidemic models are useful for understanding the general dynamics of infectious diseases, rumors, election outcomes, fads, and computer viruses [1–8]. Moreover, in the early days of emerging disease outbreaks, such as the current COVID-19 pandemic, societies rely on epidemic models for disease forecasting, as well as identifying the most effective control strategies [9–12]. To this end it is useful to quantify the risks of local epidemic outbreaks of various sizes. Within a given population, outbreak dynamics are typically described in terms of compartmental models [1,4,13]. For example, starting from some seed infection, over time individuals in a population make transitions between some number of discrete disease states (susceptible, exposed, infectious, etc.) based on prescribed probabilities for a particular disease [9,10,12,14–16]. In the limit of infinite populations the stochastic dynamics approach deterministic (mean-field) differential equations for the expected fraction of a population in each state [1,4,13,17].

Yet for real finite populations, outbreak dynamics have a wide range of different outcomes for each initial condition, which are not predicted by mean-field models. A natural and canonical question (for both statistical physics and population dynamics) is, what is the distribution of outbreak sizes? Beside stochastic simulations [1,14,17,18], methods exist for, e.g., recursively computing the outbreak statistics [16,19,20], solving the master equation for the stochastic dynamics directly by numerical linear algebra [18], or deriving scaling laws for small outbreaks near threshold [21–23]. Yet, in addition to being numerically unstable for large populations, computationally expensive, or limited in scope, such methods also fail to provide physical and analytical insights, e.g., into how unusual in extreme outbreaks occur.

Here we develop an analytical approach based on WKB methods [24–26] which provides a closed-form expression for the asymptotic outbreak distribution in SIR (susceptible-infected-recovered), SEIR (susceptible-exposed-infected-recovered), and COVID-19 models with fixed population sizes ( $N$ ) and heterogeneity in infectivity and recovery [12,27–29]. We show that each outbreak is described by a unique most-probable path, and provide an effective picture of how stochasticity is manifested during a given outbreak. For instance, compared with the expected mean-field dynamics each outbreak entails a unique depletion or boost in the pool of susceptibles and an increase or decrease in the effective recovery rate, depending on whether the final outbreak is larger or smaller than the mean-field prediction. Most importantly, unlike usual rare-event predictions for epidemic dynamics, such as extinction or other large fluctuations from an endemic state [26,30–32], and fade-out [33], our results do not rely on metastability [25,34–37] and thus are valid for the comparatively short time scales of outbreaks,  $\mathcal{O}(\ln N)$  [38]. In sharp contrast to systems undergoing escape from a metastable state, we show that the outbreak distribution corresponds to an infinite number of distinct paths—one for every possible extensive outbreak. Each outbreak connects two unique fixed points in a Hamiltonian system, *both* with nonzero probability flux. Hence, by solving a canonical problem in population dynamics and nonequilibrium statistical physics, we uncover a new degenerate class of rare events for discrete-state stochastic systems.

*Baseline model.*—We begin with the SIR model, often used as a baseline model for disease outbreaks [1,2,4,13]. Individuals are either susceptible (capable of getting infected), infected, or recovered (removed), and can make transitions between these states through two basic

processes: infection and recovery. Denoting the total number of susceptibles  $S$ , infecteds  $I$ , and recovereds  $R$  in a population of fixed size  $N$ , the probability per unit time that the number of susceptibles decreases by one and the number of infecteds increases by one is  $\beta SI/N$ , where  $\beta$  is the infectious contact rate [1,2,4]. Similarly, the probability per unit time that the number of infecteds decreases by one is  $\gamma I$ , where  $\gamma$  is the recovery rate [1,2,4]. Combining both processes results in a discrete-state system with the following stochastic reactions:



As  $N$  is assumed constant, the model is appropriate for the short time scales of early emergent-disease outbreaks, for example, with an assumed separation between the outbreak dynamics and demographic time scales, as well as reinfection [1]. From the basic reactions [Eqs. (1) and (2)], the master equation describing the probability of having  $S$  susceptibles and  $I$  infecteds at time  $t$  is

$$\begin{aligned} \frac{\partial P}{\partial t}(S, I, t) = & -\frac{\beta SI}{N}P(S, I, t) - \gamma IP(S, I, t) \\ & + \frac{\beta(S+1)(I-1)}{N}P(S+1, I-1, t) \\ & + \gamma(I+1)P(S, I+1, t). \end{aligned} \quad (3)$$

Solving this equation allows one to predict the probability that a particular proportion of a population eventually becomes infected for a given set of parameters. This is our goal here, as in many other works [1,14,16–23]. Yet, in general such equations cannot be solved analytically, and one must resort to high-dimensional numerics, recursive computations, and/or large numbers of simulations [18]. Notably, if  $N$  is large it is possible to construct an asymptotic solution to Eq. (3) for all  $\mathcal{O}(N)$  outbreaks using a WKB method [24–26], as we will show.

First, to summarize what is known for large  $N$ , let us define the fraction of individuals in each disease state  $x_w = W/N$ , where  $W \in \{S, I, R\}$ . Note that as the total population size is constant,  $1 = x_r + x_s + x_i$ . The mean-field limit of the reactions [Eqs. (1) and (2)] corresponds to a simple set of differential equations:  $\dot{x}_s = -\beta x_i x_s$ ,  $\dot{x}_i = \beta x_i x_s - \gamma x_i$ , and  $\dot{x}_r = \gamma x_i$ . Of particular interest is the total fraction of the population infected in the long-time limit,  $x_r^* = x_r(t \rightarrow \infty)$ , whose average  $\bar{x}_r^*$  can be found by integrating the mean-field system. For small initial fractions infected, the solution (according to the mean-field) depends only on the basic reproductive number  $R_0 \equiv \beta/\gamma$  [1,2,4,13] and solves the equation  $1 - \bar{x}_r^* = e^{-R_0 \bar{x}_r^*}$  [1,39].

But, what about a half, a fourth, twice, etc. of this expected outbreak, or a case in which the entire population

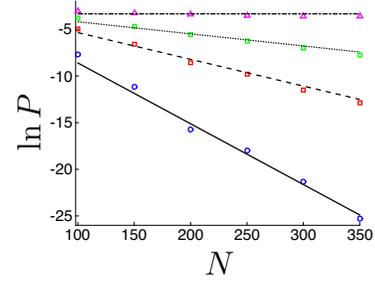


FIG. 1. Extreme outbreak probability scaling with the population size in the SIR model. Plotted is the probability that 100% (blue), 98% (red), 96% (green), and 89% (magenta) of the population are infected during an outbreak vs  $N$ . Results from  $10^{11}$  simulations (symbols) are compared with theoretical lines whose slopes are given by Eq. (7). Here  $R_0 = 2.5$ .

eventually becomes infected? Since the SIR model is inherently stochastic and governed by Eq. (3), such solutions are also possible. To get a sense of how the probabilities for various outbreaks arise, and to guide our analysis, we perform some stochastic simulations and plot (on a semilog scale) the fraction of outcomes that result in a given total-fraction infected. Examples are shown in Fig. 1 for outbreaks: 100% (blue), 98% (red), and 96% (green) when  $R_0 = 2.5$ . For reference, the mean-field outbreak of 89% (magenta) is also plotted. Here and throughout, simulations were performed using Gillespie’s direct method [1,14,40] starting from a single infectious individual. Notice that for each outbreak value,  $\ln P$  is linear in  $N$ , with a slope that depends on the outbreak,  $\ln P(x_r^*) \simeq NS(x_r^*)$ , see below. This asymptotic WKB scaling is consistent with what we expect on general theoretical grounds for large deviations in stochastic population models with a small  $\mathcal{O}(1/N)$  noise parameter [24–26,30].

Equipped with the WKB hypothesis for the distribution of outbreaks, we substitute the ansatz  $P(x_s, x_i, t) \sim \exp[-NS(x_s, x_i, t)]$  into Eq. (3) and keep leading-order terms in  $N \gg 1$ . In particular, we do a Taylor expansion of  $P(x_s, x_i, t)$ ; e.g.,  $P(x_s + 1/N, x_i - 1/N, t) \simeq e^{-NS(x_s, x_i, t) - \partial S/\partial x_s + \partial S/\partial x_i}$ . This allows one to find the leading-order solution [41], called the action, given by  $\mathcal{S}(x_s, x_i, t)$  [24,26,30]. Taking the large- $N$  limit in this way converts the master equation [Eq. (3)] into a Hamilton-Jacobi equation,  $\partial_t \mathcal{S}(x_s, x_i, t) + H(x_s, x_i, p_s, p_i) = 0$  [24,25], with a Hamiltonian given by

$$H = \beta x_i x_s (e^{p_i - p_s} - 1) + \gamma x_i (e^{-p_i} - 1). \quad (4)$$

Here the momenta of the susceptibles and infecteds are respectively defined as  $p_s = \partial \mathcal{S}/\partial x_s$  and  $p_i = \partial \mathcal{S}/\partial x_i$ .

As a consequence, in the limit of  $N \gg 1$  the outbreak dynamics satisfy Hamilton’s equations:  $\dot{x}_w = \partial H/\partial p_w$  and  $\dot{p}_w = -\partial H/\partial x_w$ , just as in analytical mechanics [42]. Furthermore, the solutions are minimum action [25] or

maximum probability. Namely, given boundary conditions for an outbreak, Hamilton's equations will provide the most likely dynamics. As in mechanics, once the dynamics are solved, the action  $\mathcal{S}(x_s, x_i, t)$  can be calculated along an outbreak path:

$$\mathcal{S}(x_s, x_i, t) = \int_0^t (p_s \dot{x}_s + p_i \dot{x}_i - H) dt'. \quad (5)$$

Before continuing our analysis, let us comment on the distribution  $P(x_s, x_i, t)$  and explain the sense in which certain outbreaks are extreme. As  $P(x_s, x_i, t)$  scales exponentially with  $N$  (for large  $N$ ), if the action  $\mathcal{S}(x_s, x_i, t)$  associated with an outbreak differs significantly from 0, the outbreak will occur with an exponentially small probability, just as we observe in Fig. 1. In fact, the special case of  $\mathcal{S} = 0$  ( $p_i = p_s = 0$ ) is nothing other than the aforementioned mean-field prediction, which nicely quantifies why it is the most likely extensive outbreak.

*Results.*—In order to find the probability distribution of outbreaks, we observe that the Hamiltonian in Eq. (4) does not depend explicitly on time; that is,  $H$  evaluated along an outbreak is conserved in time [42]. Now, we substitute  $\dot{p}_i = -\partial H / \partial x_i$  and write the Hamiltonian for the SIR model in a suggestive form  $H = -x_i \dot{p}_i$ . Thus, if we consider the same large-population limit as the usual mean-field analysis discussed above and restrict ourselves to outbreaks that start from small infection, e.g.,  $x_i(t=0) = 1/N$  with  $N \gg 1$ , it must be that  $H \simeq 0$ . Notably, because the energy is zero, we can drop the explicit time dependence in Eq. (5). As a result, since the number of infecteds grows and then decreases during the course of an outbreak with  $x_i(t) \neq 0$  for general  $t$ , one must have  $p_i = \text{const}$ .

At this point, we highlight a crucial difference between our analysis for stochastic outbreak dynamics, and the traditional use of WKB for analyzing large deviations in population models with metastable states. In the latter, the traditional  $H \simeq 0$  condition of the WKB usually derives from the fact that the model has a locally unique stable fixed point for the mean-field coordinates, e.g.,  $\dot{\mathbf{x}} = 0$  [25,26,30–32,34–37]. Common examples are stochastic switching and extinction from endemic equilibria. In our case, the zero-energy condition corresponds to a conserved momentum, and in fact, an infinite number of them. The nonzero momentum boundary conditions entailed by the conserved momenta are distinct from other known categories of extreme processes in discrete-state nonequilibrium systems and stochastic populations, and hence we uncover a new *degenerate* class.

Now that we know that outbreaks in the SIR model are defined according to a conserved momenta, i.e.,  $m \equiv e^{p_i}$ , we can equate the Hamiltonian in Eq. (4) to zero and find the nonconstant fluctuational momentum  $p_s$ , along an outbreak in terms of  $x_s$ ,  $m$ , and  $R_0$ ,

$$p_s = \ln \{R_0 x_s m^2 / [m(R_0 x_s + 1) - 1]\}. \quad (6)$$

This momentum is necessary for evaluating Eq. (5). Continuing on toward our main goal of calculating the action, we note that the integral over  $p_i$  vanishes, since it is a constant of motion and  $x_i(t=0) = x_i(t \rightarrow \infty) \simeq 0$ . Furthermore, the integral over  $H$  also vanishes since  $H \simeq 0$ . As a result, in order to determine the action, we need to compute the integral over  $p_s$  [Eq. (6)] from the initial state  $x_s = 1$  to the final state  $x_s(t \rightarrow \infty) = x_s^*$ . The only thing left for us is to express  $x_s^*$  in terms of the conserved momentum  $m$ . This can be done by using Hamilton's equations; see the Supplemental Material, Sec. I [43] for details. Doing so, we arrive at the total action accumulated in the course of an outbreak:

$$\begin{aligned} \mathcal{S}(x_s^*) &= \ln x_s^* + (1 - x_s^*) \\ &\quad \times (m(1 + R_0 x_s^*) - 1 \\ &\quad + \ln \{ [m(R_0 + 1) - 1] / (x_s^* m^2 R_0) \}). \end{aligned} \quad (7)$$

Note that  $\mathcal{S}$  is a function of  $x_s^*$  only, since for fixed  $R_0$  there is a complete mapping between the final outbreak size and  $m$  [see the Supplemental Material, Eq. (A9) [43] for  $x_s^*(m)$ ]. Equation (7) is our main result: the asymptotic solution of Eq. (3) for the distribution of all  $\mathcal{O}(N)$  outbreaks [44].

Our main result can now be tested in several ways. First, we go back to the motivating Fig. 1. Recall that our approach predicts that, as a function of  $N$ , the action gives the slope of  $\ln P(x_r^*) \simeq N\mathcal{S}(x_r^*)$ . As such, we can overlay lines in Fig. 1, where the slopes are predictions from Eq. (7). Doing so for three extreme outbreak values (as well as the mean field), we observe very good agreement, especially for larger values of  $N$ . Second, we can fix  $N$  and  $R_0$  and see how well Eq. (7) predicts the full distribution. Such comparisons with stochastic simulations are shown in the upper panel of Fig. 2(a). In particular, we plot the fraction of  $10^{12}$  simulations that resulted in an outbreak  $x_r^*$  in blue, and the solutions of Eq. (7) with a black line. Again, the agreement between the two is quite good for the population size  $N = 2000$  and  $R_0 = 1.7$ . Disagreement increases as the outbreak sizes approach  $\mathcal{O}(1/N)$ . Qualitatively, we can see that our theory captures the full cubic structure of the outbreak distribution, with local maxima at the smallest outbreak (here  $1/N$ ) and the mean-field solution,  $\bar{x}_r^* \simeq 0.69$  [24,30,33,35].

To get more insight into the outbreak distribution, one can use Eq. (7) to compute the action, e.g., in the vicinity of the mean field  $\bar{x}_r^*$ . Locally the distribution is a Gaussian around  $\bar{x}_r^*$ , with a relative variance that takes a minimum at  $R_0 \simeq 5/3$ , for which stochastic deviations from the mean-field outbreak are minimized (See the Supplemental Material, Sec. III and Fig. S1 [43] for further details on the distribution's unique shape).

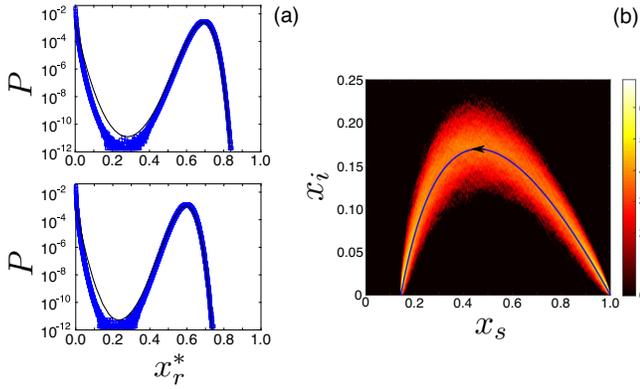


FIG. 2. Outbreak distributions. (a) Final outbreak distribution for the SIR model (top) and a COVID-19 model (bottom). Stochastic simulation results (blue squares) are compared with theory (black lines). Parameters are given in the main text. Despite the varying complexity, the outbreak distributions in both models are captured by the same theory. (b) Histogram of 2000 stochastic trajectories in the SIR model that result in the same final (non-mean-field) outbreak  $x_r^* = 0.86$ . The prediction of Eq. (9) is shown with a blue curve. Parameters are  $N = 1000$  and  $R_0 = 1.7$ . The colormap for the histogram is on natural log scale.

Before moving to more general outbreak models we mention a few important qualitative details that emerge from our approach. In particular, let us consider the stochastic dynamics for the fraction of the population infected,  $\dot{x}_i = \partial H / \partial p_i$ . Substituting Eq. (6) into  $\dot{x}_i$ , yields

$$\dot{x}_i = \beta x_i [(m-1)/(mR_0) + x_s] - (\gamma/m)x_i. \quad (8)$$

First, note that when  $m = 1$  ( $p_i = p_s = 0$ ), we uncover the mean-field SIR model system  $\dot{x}_i = \beta x_i x_s - \gamma x_i$ . From the mean field, we can recover Eq. (8) with the suggestive transformations  $x_s \rightarrow x_s + (m-1)/mR_0$ , and  $\gamma \rightarrow \gamma/m$  [45]. Recalling that each outbreak is parametrized by a unique constant  $m$ , evidently the effect of demographic stochasticity is to add an effective constant reduction (or boost) to the pool of susceptibles and to increase (or decrease) the effective recovery rate, depending on whether the final outbreak is smaller ( $m < 1$ ) or larger ( $m > 1$ ) than that of the mean field, respectively.

We can test our prediction that a conserved  $m$  constrains an entire outbreak path by picking a particular final outbreak size, corresponding to a particular value of  $m$ , and comparing to stochastic trajectories. One method for comparison is to build a histogram in the  $(x_i, x_s)$  plane from many simulations that end in the same outbreak size, and plot the constant- $m$  prediction. The latter can be found by solving the differential equation  $dx_i/dx_s = \dot{x}_i/\dot{x}_s$  from Hamilton's equations, which yields

$$x_i(x_s, m) = 1 - x_s + \ln \left[ \frac{m(R_0 x_s + 1) - 1}{m(R_0 + 1) - 1} \right] / R_0 m. \quad (9)$$

An example is shown in Fig. 2(b) for a final outbreak of 86% when  $R_0 = 1.7$  (the mean-field prediction is 69%). The colormap for the histogram is plotted along with the prediction from Eq. (9). As expected, the outbreak-path prediction lies in the maximum density region. Thus, not only does our approach predict probabilities, but also the optimal dynamics that lead to outbreaks—driven by an effective conserved momentum  $m$ .

*General model.*—We now generalize our results to more complex and realistic outbreak models. Typically, such models derive from the same basic assumptions as SIR, but have more states and free parameters. For example, epidemiological predictions for COVID-19 (at a minimum) require an incubation period of around 5 days, and an asymptomatic disease state, i.e., a group of people capable of spreading the disease without documented symptoms [27–29]. Both incubation and heterogeneity are important ingredients in a more general class of outbreak models [9,10,12,27–29]. Within this class, we assume that upon infection, susceptible individuals first become exposed, and then enter an infectious state at a finite rate  $\alpha$ . By assumption there are several possible infectious states (e.g., asymptomatic, mild, severe, tested, quarantined, etc.) that an exposed individual can enter according to prescribed probabilities [9,10,12,27–29]. In addition, infectious states can have their own characteristic infection rates and recovery times. Putting these ingredients together, let us define  $\mathcal{N}$  infectious states  $I_n$ , where  $n \in \{1, 2, \dots, \mathcal{N}\}$ , each with their own infectious contact rate  $\beta_n$  and recovery rate  $\gamma_n$ , and which appear from the exposed state with probabilities  $z_n$  [12,27–29,47]. See the Supplemental Material, Sec. II [43] for a list of reactions.

Following the WKB prescription above, the Hamiltonian for our general class of outbreak models is

$$H = \sum_n \beta_n x_{i,n} x_s (e^{p_i - p_s} - 1) + \gamma_n x_{i,n} (e^{-p_{i,n}} - 1) + \alpha z_n x_e (e^{p_{i,n} - p_e} - 1). \quad (10)$$

Despite the increased dimensionality and parameter heterogeneity, the general outbreak system defined through Eq. (10) can also be solved analytically by precisely the same approach as the baseline SIR model. As in the latter, the essential property that makes the system solvable is the constancy of all momenta except for  $p_s$ . This property ensures that, here again, there is one free constant  $m$  that determines all momenta and the final outbreak size. Demonstrating this requires a few additional steps of algebra, but the result is a simple update to Eq. (7) that involves a sum over the heterogeneities  $\{z_n, \beta_n, \gamma_n\}$ . See the Supplemental Material, Eq. (A29) [43], for a general outbreak solution. An important consequence of the general solution is that, in the special case of the SEIR model [1], where there is only one infectious state, the outbreak action is identical to the SIR model, Eq. (7). Namely, finite

incubation changes the dynamics of outbreaks, but has only a preexponential contribution to their probability.

For example, a prediction from our general analysis is shown in the lower panel of Fig. 2(a). The analytical solution (black line) is in very good agreement with stochastic simulations of a COVID-19 model with asymptomatic ( $n = 1$ ) and symptomatic ( $n = 2$ ) infectious individuals. The infection parameters [48] take realistic heterogeneous values, i.e.,  $\beta_1 = 1.8$ ,  $\beta_2 = 1.12$ ,  $\gamma_1 = 1$ ,  $\gamma_2 = 0.8$ ,  $\alpha = 2$ ,  $z_1 = 0.3$ , and  $N = 4000$  [12,27–29,47], where  $z_1 = 0.3$  is a typical value for the fraction of asymptomatic infection. Despite the increased complexity, the distribution in the more general model is also well-captured by our theory.

Before concluding, it is worth mentioning that although in real outbreaks the parameters in Eq. (10) may fluctuate in time, if the fluctuations are fast compared with outbreak time scales  $\mathcal{O}(\ln N)$  [38], we expect the distribution to approach the SIR model with effective time-averaged parameters, which can be computed using methods detailed in Refs. [49,50]. On the other hand, if the fluctuations are slow with respect to the same time scales, we expect the distribution to be described by integrating over the solution of Eq. (10), with weights given by the probability-density of rates [49,50]. In the intermediate regime, one must solve a Hamiltonian system with increased dimensionality, which includes both demographic noise and environmental variability. In this way, our results can provide a basis for understanding even more complicated outbreak dynamics.

*Conclusions.*—We solved the canonical problem of predicting the outbreak distribution of epidemics in large, fixed-sized populations. Our theory was based on the exponential scaling of the probability of extensive outbreaks on the population size, which allowed the use of a semi-classical approximation. By analyzing SIR, SEIR, and COVID-19 models, we were able to derive simple formulas for the paths and probabilities of all extensive outbreaks and find an effective picture of how stochasticity is manifested during outbreaks. Most importantly we showed that, unlike other well-known examples of rare events in population models, the statistics of extreme outbreaks depend on an infinite number of minimum-action paths satisfying a unique set of boundary conditions with conserved momenta. Due to their distinct and degenerate phase-space topology, extreme outbreaks represent a new class of rare processes for discrete-state stochastic systems. As with other extreme processes, our solution can form the basis for predictions in many other scenarios, including stochastic outbreaks mediated through complex networks.

J. H. and I. B. S. were supported by the U.S. Naval Research Laboratory funding (N0001419WX00055), and the Office of Naval Research (N0001419WX01166 and N0001419WX01322). M. A. was supported by the Israel Science Foundation Grant No. 531/20, and by the Humboldt Research Fellowship for Experienced Researchers of the Alexander von Humboldt Foundation.

- [1] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ, 2008).
- [2] H. Andersson and T. Britton, *Stochastic Epidemic Models and their Statistical Analysis* (Springer-Verlag, New York, 2000).
- [3] R. Pastor-Satorras, C. Castellano, P. Van, Miegheem, and A. Vespignani, *Rev. Mod. Phys.* **87**, 925 (2015).
- [4] H. W. Hethcote, *SIAM Rev.* **42**, 599 (2000).
- [5] L. Billings, W. M. Spears, and I. B. Schwartz, *Phys. Lett. A* **297**, 261 (2002).
- [6] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [7] J. Hindes and M. Assaf, *Phys. Rev. Lett.* **123**, 068301 (2019).
- [8] A. Volkening, D. F. Linder, M. A. Porter, and G. A. Rempala, *SIAM Rev.* **62**, 837 (2020).
- [9] R. C. Reiner *et al.*, *Nat. Med.* **27**, 94 (2021).
- [10] E. L. Ray *et al.*, medRxiv (2020), 10.1101/2020.08.19.20177493.
- [11] A. Vazquez, *Phys. Rev. E* **103**, L030301 (2021).
- [12] A. Catching, S. Capponi, M. T. Yeh, S. Bianco, and R. Andino, medRxiv (2021), 10.1101/2020.08.12.20173047.
- [13] F. Brauer, *Infect. Dis. Modell.* **2**, 113 (2017).
- [14] T. Ganyani, C. Faes, and N. Hens, *Annu. Rev. Stat. Appl.* **8**, 69 (2021).
- [15] C. M. Batistela, D. P. F. Correa, Á. M. Bueno, and J. C. Piqueira, *Chaos, Solitons, Fractals* **142**, 110388 (2021).
- [16] J. C. Miller, arXiv:1907.05138.
- [17] L. J. S. Allen, *Infect. Dis. Modell.* **2**, 128 (2017).
- [18] T. House, J. V. Ross, and D. Sirl, *Proc. R. Soc. A* **469**, 20120436 (2013).
- [19] F. Ball, *Adv. Appl. Probab.* **18**, 289 (1986).
- [20] F. Ball and D. Clancy, *Adv. Appl. Probab.* **25**, 721 (1993).
- [21] E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **69**, 050901 (R) (2004).
- [22] E. Ben-Naim and P. L. Krapivsky, *Eur. Phys. J. B* **85**, 145 (2012).
- [23] S. Singh and C. R. Myers, *Phys. Rev. E* **89**, 042108 (2014).
- [24] M. Assaf and B. Meerson, *J. Phys. A* **50**, 263001 (2017).
- [25] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, *J. Chem. Phys.* **100**, 5735 (1994).
- [26] O. Ovaskainen and B. Meerson, *Trends Ecol. Evol.* **25**, 643 (2010).
- [27] R. Subramanian, Q. He, and M. Pascual, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019716118 (2021).
- [28] C. C. Kerr *et al.*, medRxiv (2020), 10.1101/2020.05.10.20097469.
- [29] I. B. Schwartz, J. H. Kaufman, K. Hu, and S. Bianco, medRxiv (2020), 10.1101/2020.04.16.20068387.
- [30] M. Assaf and B. Meerson, *Phys. Rev. E* **81**, 021116 (2010).
- [31] J. Hindes and I. B. Schwartz, *Phys. Rev. Lett.* **117**, 028302 (2016).
- [32] A. J. Black and A. J. McKane, *J. Stat. Mech.* (2011) P12006.
- [33] B. Meerson and P. V. Sasorov, *Phys. Rev. E* **80**, 041130 (2009).
- [34] I. Nåsell, *Extinction and Quasi-Stationarity in the Stochastic Logistic SIS Model* (Springer, New York, 2011).

- [35] I. B. Schwartz, E. Forgoston, S. Bianco, and L. B. Shaw, *J. R. Soc. Interface* **8**, 1699 (2011).
- [36] A. Kamenev and B. Meerson, *Phys. Rev. E* **77**, 061107 (2008).
- [37] E. Forgoston and R. O. Moore, *SIAM Rev.* **60**, 969 (2018).
- [38] M. Turkyilmazoglu, *Physica (Amsterdam)* **422D**, 132902 (2021).
- [39] T. Harko, F. S. Lobo, and M. Mak, *Appl. Math. Comput.* **236**, 184 (2014).
- [40] D. T. Gillespie, A. Hellander, and L. R. Petzold, *J. Chem. Phys.* **138**, 170901 (2013).
- [41] Sub-leading order contributions to the probability can be found by continuing the large-N expansion [24,26,30,32].
- [42] L. D. Landau and E. M. Lifshitz, *Mechanics, Third Edition: Volume 1 (Course of Theoretical Physics)*, 3rd ed. (Butterworth-Heinemann, London, 1976).
- [43] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.128.078301> for more details on the model and additional results from the analysis and simulations.
- [44] For brevity we have dropped the dependence on  $x_i$  in the final-outbreak action given by Eq. (7), since all final states have the same  $x_i = 0$ .
- [45] A similar effect occurs in cell biology in a mRNA-protein genetic circuit, where fluctuations in the mRNA copy number can be effectively accounted for by taking a protein-only model with a modified production rate [46].
- [46] E. Roberts, S. Be'er, C. Bohrer, R. Sharma, and M. Assaf, *Phys. Rev. E* **92**, 062717 (2015).
- [47] J. Hindes, S. Bianco, and I. B. Schwartz, *PLoS One* **16**, e0244706 (2021).
- [48] For COVID-19 modelling, a typical choice for time units would be  $t = 1$  corresponding to 10 days.
- [49] M. Assaf, A. Kamenev, and B. Meerson, *Phys. Rev. E* **78**, 041123 (2008).
- [50] M. Assaf, E. Roberts, Z. Luthey-Schulten, and N. Goldenfeld, *Phys. Rev. Lett.* **111**, 058102 (2013).