

Large deviations in non-Markovian stochastic epidemics

Matan Shmunik  and Michael Assaf **Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel*

(Received 11 September 2025; accepted 25 April 2026; published 1 June 2026)

We develop a framework for non-Markovian, well-mixed susceptible-infected-recovered (SIR) and susceptible-infected-susceptible (SIS) models beyond mean field, utilizing the continuous-time random walk formalism. Using a gamma distribution for the infection and recovery interevent times as a test case, we derive asymptotical late-time master equations with effective memory kernels and obtain analytical predictions for the final outbreak-size distribution in the SIR model, and quasistationary distribution and disease lifetime in the SIS model. In particular, we show that varying the width of the interevent time distribution can greatly alter the outbreak-size distribution and disease lifetime. Overall, our analysis, confirmed against numerical simulations, paves the way for studying large deviations in structured populations on degree-heterogeneous networks.

DOI: [10.1103/bswp-wrfc](https://doi.org/10.1103/bswp-wrfc)

Introduction. The challenge of modeling and predicting pandemics has motivated extensive development of deterministic and stochastic epidemiological models [1–8]. Their importance was underscored during COVID-19 [9,10], when forecasts guided policies worldwide. These models typically divide populations into disease-status compartments. The susceptible-infected-susceptible (SIS) model [11] with susceptible and infected compartments describes infections with short-lived immunity (e.g., influenza or the common cold), while the susceptible-infected-recovered (SIR) model [8] adds a recovered compartment to capture long-term immunity, which is relevant for diseases such as measles, smallpox, polio, and COVID-19. Generalized models include reinfection (SIRS), an additional exposed compartment [6], time-varying rates [12,13], or structured populations, where network-based approaches [14–20] capture heterogeneity in population structure. Mean-field quantities have been computed for both models: the mean outbreak size and epidemic threshold in the SIR model, and endemic state and relaxation dynamics in the SIS model [3,5,11,14]. Stochastic dynamics have also been explored, including the outbreak-size distribution for the SIR model and quasistationary distribution (QSD) and disease mean time to extinction (MTE) for the SIS model [13,21–31].

A key limitation of most approaches is the Markovian assumption [7,32,33], which states that the system dynamics depend solely on the present state and are independent of prior history. Under this assumption, both infection and recovery periods are exponentially distributed. However, empirical evidence suggests that transmission and recovery rates are often explicitly time dependent, and consequently, the

corresponding waiting-time (WT) distributions are more accurately described by nonexponential distributions, such as log-normal, gamma, or Weibull [34–36]. As a result, infection and recovery dynamics are generally non-Markovian, rendering analytical frameworks based on Markovian assumptions often inadequate for describing such processes.

Most studies investigating non-Markovian epidemic processes on various network topologies have focused on the mean-field aspects of the dynamics [33,37–45]. These studies have shown that, even when average rates are unchanged, non-exponential infection and recovery times can drastically affect disease spread, prevalence, and overall epidemic behavior. Conversely, certain features of non-Markovian dynamics can still be approximated by rescaling Markovian rates (see, e.g., Refs. [45,46]). In this context, Boguñá *et al.* [39] proposed a generalized Gillespie algorithm applicable to nonexponential infection and recovery processes, which was later refined into the more efficient Laplace Gillespie method by Masuda and Rocha [44]. Notably, while the role of noise in non-Markovian epidemics has also been studied in both SIR [47–50] and SIS [37] models, a systematic analysis of the interplay between non-Markovianity and demographic noise on large deviations in these models remains lacking.

Recent progress has extended non-Markovian frameworks to chemical and biological systems, using the continuous-time random walk formalism with nonexponential WTs [51]. This was later adapted in Refs. [52–54] to models of gene expression, population dynamics, and competition. Here, we combine this framework with the WKB approach [23,26,55] of Hindes *et al.* [28] to study non-Markovian epidemic dynamics beyond mean field in a well-mixed setting. We calculate the memory kernels for general WTs and derive the effective late-time master equation [51,53]. We then compute the mean outbreak size and its full distribution within the SIR model, using gamma-distributed WTs as a prototypical example. We also analyze the SIS model, deriving the metastable mean, its full distribution (QSD), and MTE—the first passage time to the absorbing, disease-free state [24,56]. Finally, we

*Contact author: michael.assaf@mail.huji.ac.il

test our formalism using empirical WTs for both infection and recovery.

It is important to note that in this work we focus on well-mixed populations, i.e., fully connected networks, in which all individuals are considered identical. As such, for simplicity, we consider only two global reactions: infection and recovery, both with generic WT distributions, whose mean depends on the system's state. While this constitutes an oversimplified representation of realistic scenarios, it turns out that our model can reproduce, at least qualitatively, key features of large deviations in non-Markovian epidemic dynamics on complex networks.

Non-Markovian SIR model. Within the well-mixed SIR model, given that the rates of infection and recovery per individual are β and γ , respectively, the discrete-state stochastic reactions can be written as $(S, I) \rightarrow (S - 1, I + 1)$ with mean rate $\beta SI/N$, and $(I, R) \rightarrow (I - 1, R + 1)$ with mean rate γI . Defining the population fractions as $x_s = S/N$, $x_i = I/N$, and $x_r = R/N$, assuming $N \gg 1$ and ignoring noise, the mean-field rate equations read

$$\dot{x}_s = -\beta x_s x_i, \quad \dot{x}_i = \beta x_s x_i - \gamma x_i, \quad \dot{x}_r = \gamma x_i, \quad (1)$$

such that $x_s + x_i + x_r = 1$. Here, the underlying assumption is that reactions occur with exponential WTs. Yet, in reality, the WT distributions, henceforth denoted by $\psi_1(t)$ for infection and $\psi_2(t)$ for recovery, are not necessarily exponential.

Initially, we focus on the case of non-Markovian, gamma-distributed infection and Markovian recovery [38,39], while other WT choices are discussed in the Supplemental Material [57]. The WT distributions are

$$\psi_1(t) = (\alpha \lambda_1)^\alpha t^{\alpha-1} e^{-\alpha \lambda_1 t} / \Gamma(\alpha), \quad \psi_2(t) = \lambda_2 e^{-\lambda_2 t}, \quad (2)$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ is the gamma function. Notably, the definition of the gamma distribution here is such that the shape parameter α does not affect the mean. The mean rates are set to equal the Markovian counterparts, $\lambda_1 = NR_0 x_s x_i$ and $\lambda_2 = N x_i$, respectively, where $R_0 = \beta/\gamma$ is the basic reproductive number. The shape parameter α controls the tail of the WT distribution and, as such, can be effectively viewed as a form of memory, where lower values of α entail stronger memory (broader distribution) while higher values of α entail weaker memory (narrower distribution).

To determine the outbreak-size distribution within the SIR model, we write the corresponding non-Markovian master equation for the probability $P_{S,I}$ to find S susceptibles and I infected at time t , which reads [52,53]

$$\begin{aligned} \frac{\partial P_{S,I}}{\partial t} = & \int_0^t [M_1(S+1, I-1, t-t') P_{S+1, I-1}(t') \\ & - M_1(S, I, t-t') P_{S,I}(t') \\ & + M_2(S, I+1, t-t') P_{S, I+1}(t') \\ & - M_2(S, I, t-t') P_{S,I}(t')] dt'. \end{aligned} \quad (3)$$

In both infection and recovery, the dynamics depend on the full time history via the memory kernels M_i . The latter can be found by Laplace transforming the master equation and using the specific structure of the WT distributions, $\psi_i(t)$ (see Sec. A of the Supplemental Material [57]). Even though the SIR dynamics have no metastability and include an

infectious wave that quickly dies out, to explore the final outbreak distribution it suffices to look at the asymptotic late-time dependence of M_i in the limit of $t \rightarrow \infty$ [58]. Below, we will also provide numerical justification for this claim. Thus, we use the final value theorem: $\lim_{t \rightarrow \infty} f(t) = \lim_{u \rightarrow 0} u \tilde{f}(u)$, for the memory kernel functions, where u is the Laplace variable, and $\tilde{f}(u)$ is the Laplace transform of $f(t)$. Defining the normalized asymptotic memory kernels: $\mathcal{M}_i := (1/N) \lim_{t \rightarrow \infty} M_i(t) = (1/N) \lim_{u \rightarrow 0} u \tilde{M}_i(u)$ (see details in Sec. A of the Supplemental Material [57]) and using Eq. (3), we obtain the effective late-time master equation for the SIR model:

$$\begin{aligned} \frac{\partial P_{S,I}}{\partial t} = & N[\mathcal{M}_1(S+1, I-1) P_{S+1, I-1} - \mathcal{M}_1(S, I) P_{S,I} \\ & + \mathcal{M}_2(S, I+1) P_{S, I+1} - \mathcal{M}_2(S, I) P_{S,I}], \end{aligned} \quad (4)$$

$$\text{with } \mathcal{M}_1 = \frac{x_i}{[1 + 1/(R_0 x_s \alpha)]^\alpha - 1}, \quad \mathcal{M}_2 = x_i. \quad (5)$$

Here, $\mathcal{M}_1$ and $\mathcal{M}_2$ are the effective infection and recovery rates under non-Markovian (gamma-distributed) infection and Markovian recovery. For $\alpha = 1$ (exponential infection), we recover the Markovian rate $\mathcal{M}_1 = R_0 x_s x_i$, whereas decreasing α increases the infection rate and greatly alters the dynamics. Notably, $\mathcal{M}_1$ and $\mathcal{M}_2$ can also be found for other choices of WTs (see Sec. B of the Supplemental Material [57]).

Having found the effective rates, we now use the WKB approach as in Ref. [28] to compute the final outbreak-size distribution. Substituting $P_{S,I} \sim e^{-N \mathcal{S}(x_s, x_i, t)}$ into Eq. (4) yields, in the leading order in $N \gg 1$ [59], a Hamilton-Jacobi equation $\partial \mathcal{S}(x_s, x_i, t) / \partial t + \mathcal{H}(x_s, x_i) = 0$, where $\mathcal{S}(x_s, x_i, t)$ is the action function, with the Hamiltonian

$$\mathcal{H} \equiv \mathcal{M}_1(x_s, x_i)(e^{p_i - p_s} - 1) + \mathcal{M}_2(x_s, x_i)(e^{-p_i} - 1). \quad (6)$$

Here, we have defined the momenta $p_i = \partial \mathcal{S} / \partial x_i$, $p_s = \partial \mathcal{S} / \partial x_s$ in analogy to classical mechanics. In Sec. C of the Supplemental Material [57], we show that p_i is a constant of motion and use Hamilton's equations to derive a relation between p_i and final susceptible fraction $x_s^* = x_s(t \rightarrow \infty)$. Defining $m = e^{p_i}$, we find an implicit relation between m and x_s^* :

$$\int_{x_s^*}^1 [m(\mathcal{M}_1/\mathcal{M}_2 + 1) - 1]^{-1} dx_s = \int_0^{1-x_s^*} dx_r. \quad (7)$$

This allows to compute the action function $\mathcal{S}(x_s, x_i, t) = \int_0^t (p_s \dot{x}_s + p_i \dot{x}_i - \mathcal{H}) dt'$ [28] (see Sec. C of the Supplemental Material [57]), which reads

$$\mathcal{S}(x_s^*) = \int_1^{x_s^*} \ln\{m^2 \mathcal{M}_1 / [m(\mathcal{M}_1 + \mathcal{M}_2) - \mathcal{M}_2]\} dx_s, \quad (8)$$

and is a function of x_s^* only via Eq. (7). Finally, the final outbreak-size distribution is given by $P(x_s^*) \sim e^{-N \mathcal{S}(x_s^*)}$.

To quantify outbreak variability, we compute the standard deviation by expanding the action $\mathcal{S}$ to second order around the mean final susceptible fraction $\bar{x}_s^*$, obtained at $p_s = p_i = 0$. Plugging this in Eq. (7), we get $\int_{\bar{x}_s^*}^1 (\mathcal{M}_2/\mathcal{M}_1) dx_s = \int_0^{1-\bar{x}_s^*} dx_r$, an implicit equation for $\bar{x}_s^*$. Using $\mathcal{M}_i$ from Eq. (5)

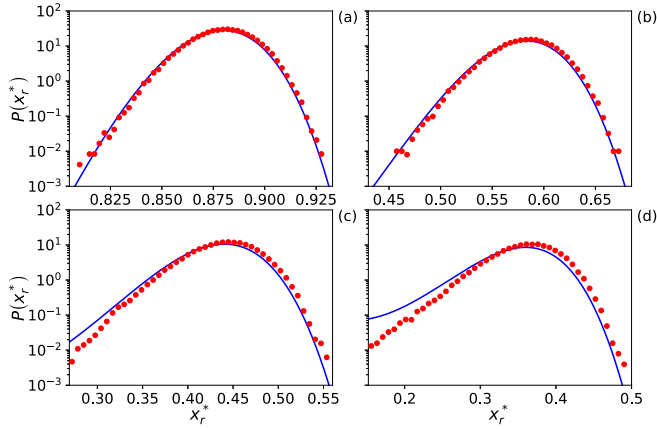


FIG. 1. Outbreak-size distributions for the SIR model, with $\alpha = 0.5, 1, 1.5$, and 2 in panels (a)–(d), respectively, for gamma-distributed infection and Markovian recovery, $N = 5000$, $R_0 = 1.5$, and 10^5 runs per α . Simulations (symbols) are compared with theory (solid line). In all panels $I(0) = 1$, and simulations resulting in outbreak fractions $< 10^{-2}$ were omitted [61].

yields $1 - \bar{x}_s^* = (2\alpha R_0)^{-1} f(\bar{x}_s^*)$, with

$$f(\bar{x}_s^*) = B\left(\frac{\alpha R_0}{1 + \alpha R_0}; 1 - \alpha, -1\right) - B\left(\frac{\alpha R_0 \bar{x}_s^*}{1 + \alpha R_0 \bar{x}_s^*}; 1 - \alpha, -1\right), \quad (9)$$

where $B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$ is the incomplete beta function. Here, for $\alpha = 1$, we recover the known mean-field result of $1 - \bar{x}_s^* = -\ln \bar{x}_s^*/R_0$ [8].

Using Eq. (8) for the action $\mathcal{S}$, we can derive the standard deviation σ by performing a Gaussian approximation on $P(x_s^*)$ around $\bar{x}_s^*$, yielding $\sigma \simeq |N\mathcal{S}''(\bar{x}_s^*)|^{-1/2}$ [28]. Defining $\xi(\bar{x}_s^*) = \mathcal{M}_2(\bar{x}_s^*)/\mathcal{M}_1(\bar{x}_s^*)$ and $I(\bar{x}_s^*) = \int_{\bar{x}_s^*}^1 \xi(z)^2 dz$, we find (see Sec. C of the Supplemental Material [57])

$$\mathcal{S}''(x_s^*)|_{\bar{x}_s^*} = [\xi(\bar{x}_s^*) - 1]^2 / [1 - \bar{x}_s^* + I(\bar{x}_s^*)], \quad (10)$$

where using Eq. (5), $\xi(\bar{x}_s^*) = [1 + 1/(R_0 \alpha \bar{x}_s^*)]^\alpha - 1$ and $I(\bar{x}_s^*)$ can be expressed by beta functions.

To validate our theory, we ran simulations using the next reaction method with simplified WT sampling [60], see Ref. [61], which in this case is accurate and faster compared to the algorithms in Refs. [39,44]. Figures 1 and 2 show results for the non-Markovian SIR model with gamma-distributed infection and exponential recovery. In Fig. 1, we show examples of the final outbreak-size distribution for four different α 's [62]. Here, theory and simulations agree well as long as the action is large [24,26,28,55]. In Fig. 2, we show the mean outbreak fraction and its variance as function of α . As α increases, typical infection times become longer, and the mean drops, while the standard deviation increases. We also see that the WKB approximation loses its accuracy as the mean outbreak size approaches zero.

Non-Markovian SIS model. Here, a susceptible individual can get infected at a rate β and infected individuals recover at a rate γ . The Markovian mean-field dynamics satisfy $\dot{x}_i = \beta x_s x_i - \gamma x_i$, where $x_i + x_s = 1$. Notably, the late-time

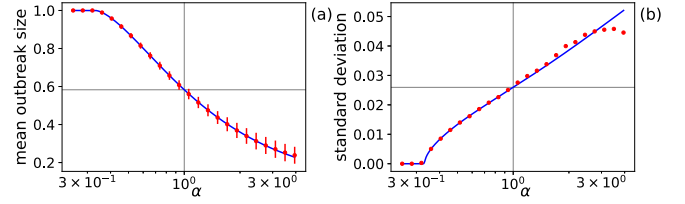


FIG. 2. Mean outbreak-size distribution x_r^* [panel (a)] and the distribution's standard deviation σ [panel (b)] vs α . Here, infection is gamma distributed and recovery is exponential, $N = 5000$, $R_0 = 1.5$, and we ran 10^4 realizations per α . Simulations (symbols) are compared with theory (solid line). In all panels $I(0) = 1$, and simulations resulting in outbreak fractions $< 10^{-2}$ were omitted [61].

dynamics of this model are markedly different from the SIR model. Here, for $N \gg 1$, the system enters a long-lived metastable state, which then slowly decays due to an exponentially small probability flux into the absorbing state at $I = 0$. As we are interested in finding the statistics of the metastable state and MTE, we again use the long-time effective memory kernels defined above and arrive at the late-time master equation for $P_I(t)$ —the probability of finding I infected at time t :

$$\partial P_I / \partial t = N[\mathcal{M}_1(I-1)P_{I-1} - \mathcal{M}_1(I)P_I + \mathcal{M}_2(I+1)P_{I+1} - \mathcal{M}_2(I)P_I]. \quad (11)$$

This master equation is valid after the system has relaxed to the long-lived metastable endemic state, and the memory kernels $\mathcal{M}_i$ are given by Eq. (5) with $x_s = 1 - x_i$.

To quantify the effect of non-Markovian infection on the endemic state x_i^* , we multiply Eq. (11) by I and sum over all I . This yields the modified rate equation $\dot{x}_i = \mathcal{M}_1(x_i) - \mathcal{M}_2(x_i)$, with x_i^* found by solving $\mathcal{M}_1(x_i^*) = \mathcal{M}_2(x_i^*)$. Using Eq. (5) with $x_s = 1 - x_i$, yields

$$x_i^* = 1 - [\alpha R_0 (2^{1/\alpha} - 1)]^{-1}. \quad (12)$$

At $\alpha = 1$ (exponential WTs), we recover the Markovian result $x_i^* = 1 - 1/R_0$. However, at $\alpha \neq 1$, one can define a new effective R_0 , $R_0^{\text{eff}} = \alpha R_0 (2^{1/\alpha} - 1)$, such that $x_i^* = 1 - 1/R_0^{\text{eff}}$. This means that, in order to reproduce the effect of non-Markovian infection, one can instead take Markovian infection and change the infection rate $\beta \rightarrow \beta \alpha (2^{1/\alpha} - 1)$ such that at $\alpha \rightarrow 0$, the Markovian infection rate should tend to ∞ , while at $\alpha \rightarrow \infty$, the infection rate should tend to $\beta \ln 2$. Consequently, we find that at $\alpha < 1$ an endemic state can persist even when each infected transmits to fewer than one other individual, or in other words, a subcritical epidemic with $R_0 < 1$ may still reach a nonzero endemic state. Note that the relaxation time of the dynamics $\tau_r = [\mathcal{M}_2'(x_i^*) - \mathcal{M}_1'(x_i^*)]^{-1}$ can also be computed. Using Eq. (5) we obtain $\tau_r = 2^{1/\alpha-1} / \{\alpha(2^{1/\alpha} - 1)[R_0 \alpha (2^{1/\alpha} - 1) - 1]\}$, which coincides with the known result of $\tau_r = 1/(R_0 - 1)$ at $\alpha = 1$.

We now analyze the effective master equation (11) to determine the QSD of the metastable state and the MTE under non-Markovian infection. Assuming the metastable state decays slowly, we set $P_I(t) = \pi(x_i) e^{-t/\tau_{\text{ext}}}$, where $\pi(x_i)$ is the QSD and τ_{ext} is the exponentially large MTE. We now apply the WKB approximation, $\pi(x_i) \sim e^{-N\mathcal{S}(x_i)}$, where $\mathcal{S}(x_i)$ is the

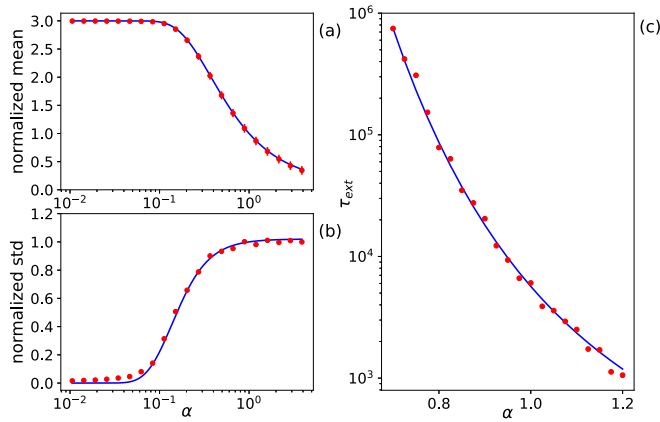


FIG. 3. In panels (a) and (b) we show the normalized mean x_i^* and normalized standard deviation σ vs α , for gamma-distributed infection and Markovian recovery. Here, $N = 1000$, $R_0 = 1.5$, and we ran 10^3 realizations per α . Simulations (symbols) are compared with theory (line). In panel (c), we plot the MTE vs α for the same WT, $N = 110$, $R_0 = 1.51$, and 10^2 runs per α . Simulations (symbols) are compared with theory (line) (see text). In panels (a)–(c), $x_i(0) = x_i^*$.

action function. Expanding to leading order in $N \gg 1$ and using Eq. (5), we find $S'(x_i) = \ln[\mathcal{M}_2(x_i)/\mathcal{M}_1(x_i)] = \ln[1 + 1/(R_0\alpha(1 - x_i))^\alpha - 1]$, which provides the QSD, $\pi(x_i)$:

$$\pi(x_i) \sim e^{-N \int_{x_i^*}^{x_i} \ln[\mathcal{M}_2(x')/\mathcal{M}_1(x')] dx'} \quad (13)$$

The QSD can be explicitly written using Eq. (5) in terms of elementary functions, for each value of α , and generalizes the result for Markovian reactions [21,24]. In Fig. S1 of the Supplemental Material [57], we show excellent agreement between simulated QSDs for various values of α and our theoretical prediction (see details in Sec. D of the Supplemental Material [57]).

We can also compute the QSD's variance, σ^2 , via a Gaussian approximation around x_i^* [24]. This yields

$$\sigma^2 = \frac{2^{1/\alpha-1}}{NR_0\alpha^2(2^{1/\alpha} - 1)^2} = \frac{1 - x_i^*}{2\alpha N(1 - 2^{-1/\alpha})}, \quad (14)$$

where we have used Eq. (5). At $\alpha = 1$, the Markovian result is recovered, $\sigma^2 = 1/(NR_0)$. As α increases, typical fluctuations can greatly increase, making disease clearance more likely. In Figs. 3(a) and 3(b), we show how the shape parameter α affects the metastable mean (12) and its standard deviation (14), both normalized by their Markovian values [$x_i^*(\alpha = 1) = 1/3$, $\sigma(\alpha = 1) \simeq 0.026$, see caption]. As α increases, the mean decreases drastically [38,39], and the variance increases; this occurs since the typical time between infection events gets longer as the WT distribution becomes narrower. At $\alpha \rightarrow 0$, the times between infection events are very short and the mean approaches 1, while the variance approaches 0.

The dependence of the variance on α as observed in Fig. 3(b) can be explained by looking at the right-hand side of Eq. (14), where the standard deviation is expressed as a function of the mean x_i^* and α (instead of R_0 and α) [63]. This reveals two effects shaping σ . The first and strongest effect is the monotone relation $\sigma \sim \sqrt{1 - x_i^*}$, dominating the

behavior. The second effect stems from non-Markovianity, $\sigma \sim [2\alpha(1 - 2^{-1/\alpha})]^{-1/2}$; for non-Markovian infection, it partially opposes the main effect of the increase in σ as α is increased; yet, it reinforces the main effect in the case of non-Markovian recovery (see Sec. E of the Supplemental Material [57]). Notably, as x_i^* approaches 0, the WKB approximation is expected to break down as the action is no longer large. This effect is more evident for non-Markovian recovery (see Sec. E of the Supplemental Material [57]).

Finally, having computed the complete QSD (13), we can evaluate the MTE, using $\tau_{\text{ext}} \sim e^{NS(0)}$, where $S(0)$ is the action barrier to extinction [24,26,55,64]. The integral in Eq. (13) from x_i^* to 0 yields the MTE and can be explicitly computed for a given α . In particular, a simplified expression can be obtained, at $|\alpha - 1| \ll 1$, which can provide insight into how the disease lifetime changes as one deviates from Markovianity. In the leading $|\alpha - 1| \ll 1$ order, we obtain $S(0) \simeq S_0 - [(\alpha - 1)/(2R_0)][(R_0 + 1)^2 \ln(1 + R_0^{-1}) - R_0 + 1 + \ln(R_0/16)]$, where $S_0 = \ln R_0 + R_0^{-1} - 1$ is the action barrier in the $\alpha = 1$ case [21,23,24,26]. In Fig. 3(c), we plot the MTE versus α for gamma-distributed infection and exponential recovery. Very good agreement holds as long as the action is large (i.e., $\ln \tau_{\text{ext}} \gg 1$), which is the case for the entire range of the figure. Here, the solid line shows the theoretical result, $\tau_{\text{ext}} \simeq A(\alpha, R_0, N)e^{NS(0)}$; it includes an additional preexponential factor $A(\alpha, R_0, N)$ [24,26], numerically fitted (for fixed N and R_0) to be $A \sim \alpha^2$.

Discussion. We have studied how non-Markovian reactions shape long-term epidemic dynamics, by deriving the late-time asymptotics of the master equation, and memory kernels emanating from the WT. We have shown, within the SIR model, that under non-Markovian infection and exponential recovery, the mean outbreak size and its entire distribution strongly depend on the shape parameter α ; notably, as α grows, the outbreak risk greatly diminishes. Within the SIS model, we have shown that as α increases, disease prevalence decreases [38,39], and the risk of disease eradication greatly increases.

Having focused so far on non-Markovian infection, we now show the model's applicability to realistic scenarios with both infection and recovery being non-Markovian. Empirical studies of acute infections like influenza and COVID-19 consistently report nonexponential (gamma, Weibull, or log-normal) WT for infection and recovery, typically with gamma shape parameters $1 < \alpha_{\text{inf}}, \alpha_{\text{rec}} < 6$ [65–74]. Infection WT typically measure the time from inferred exposure or symptom onset to secondary transmission, while recovery WT correspond to infectious periods or time from a positive test to a negative one. Population-level heterogeneity further shapes these distributions. For example, it is expected that elderly, immunocompromised, or high-contact individuals will show longer, more variable recovery times and lower shape parameters, whereas children and young adults in community settings will tend to progress and recover more uniformly [69,71,74].

In Fig. 4, we show the dramatic impact of incorporating non-Markovian reactions. Here, numerical and theoretical heatmaps of the outbreak-size distribution's coefficient of variation (COV) are compared for gamma-distributed infection and recovery, and are plotted versus the WT's shape

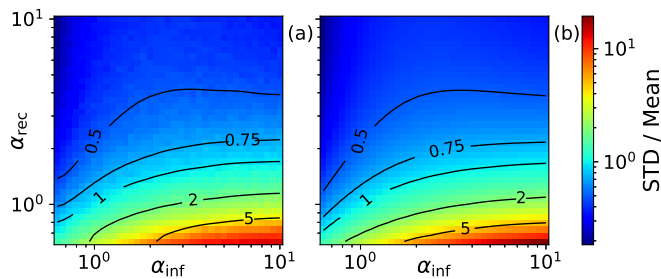


FIG. 4. Numerical (left) and theoretical (right) heatmaps of the outbreak-size distribution COV (normalized by the Markovian COV) vs the infection and recovery shape parameters, for a well-mixed system with $N = 5000$, $R_0 = 2$, and 10^3 runs per point. Black lines are equi-COV contour lines.

parameters in a well-mixed setting. Notably, the COV, indicating the relative error in the expected outbreak size, can greatly exceed the Markovian prediction, especially in regions of large values of the infection shape parameter, indicating synchronized and rapid progression from exposure to infectiousness.

Future research should aim at extending our formalism to realistic non-Markovian epidemic settings involving *structured populations* residing on complex heterogeneous networks. The mean-field aspects of non-Markovian epidemics have already been studied on such networks, and the next challenge is to study the interplay between large deviations and complex network topology under non-Markovian dynamics. Therefore, upon successful extension of the theory to real-life network topologies, and accounting for empirically grounded shape parameters for both infection and recovery (as done in Fig. 4), this framework is expected to provide a more accurate representation of epidemic dynamics. In particular, it is expected to improve estimates of epidemic thresholds, outbreak sizes, disease clearance times, and, most importantly, intervention measures.

Acknowledgment. The authors wish to thank Ami Taitelbaum for many useful discussions.

Data availability. The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

- [1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, New York, 1991).
- [2] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. London Ser. A* **115**, 700 (1927).
- [3] L. J. S. Allen, Some discrete-time *SI*, *SIR*, and *SIS* epidemic models, *Math. Biosci.* **124**, 83 (1994).
- [4] D. Mollison, *Epidemic models: Their Structure and Relation to Data* (Cambridge University Press, 1995).
- [5] D. J. Daley and J. M. Gani, *Epidemic Modelling: An Introduction* (Cambridge University Press, Cambridge, UK, 1999).
- [6] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* **42**, 599 (2000).
- [7] M. J. Keeling and K. T. Eames, Networks and epidemic models, *J. R. Soc. Interface* **2**, 295 (2005).
- [8] J. Neipel, J. Bauermann, S. Bo, T. Harmon, and F. Jülicher, Power-law population heterogeneity governs epidemic waves, *PLoS One* **15**, e0239678 (2020).
- [9] W. Yang, D. Zhang, L. Peng, C. Zhuge, and L. Hong, Rational evaluation of various epidemic models based on the COVID-19 data of China, *Epidemics* **37**, 100501 (2021).
- [10] R. C. Reiner, Jr., R. M. Barber, J. K. Collins, P. Zheng, C. Adolph, J. Albright, C. M. Antony, A. Y. Aravkin, S. D. Bachmeier, B. Bang-Jensen, *et al.*, Modeling COVID-19 scenarios for the United States, *Nat. Med.* **27**, 94 (2021).
- [11] H. W. Hethcote, Three basic epidemiological models, in *Applied Mathematical Ecology*, edited by S. A. Levin, T. G. Hallam, and L. J. Gross (Springer, Berlin, 1989), pp. 119–144.
- [12] J. L. Aron and I. B. Schwartz, Seasonality and period-doubling bifurcations in an epidemic model, *J. Theor. Biol.* **110**, 665 (1984).
- [13] L. B. Shaw and I. B. Schwartz, Fluctuating epidemics on adaptive networks, *Phys. Rev. E* **77**, 066101 (2008).
- [14] R. Pastor-Satorras, C. Castellano, P. van Mieghem, and A. Vespignani, Epidemic processes in complex networks, *Rev. Mod. Phys.* **87**, 925 (2015).
- [15] B. Karrer and M. E. Newman, Message passing approach for general epidemic models, *Phys. Rev. E* **82**, 016101 (2010).
- [16] J. Lindquist, J. Ma, P. van den Driessche, and F. H. Willeboordse, Effective degree network disease models, *J. Math. Biol.* **62**, 143 (2011).
- [17] J. C. Miller, A. C. Slim, and E. M. Volz, Edge-based compartmental modelling for infectious disease spread, *J. R. Soc. Interface* **9**, 890 (2012).
- [18] M. J. Keeling, The effects of local spatial structure on epidemiological invasions, *Proc. R. Soc. London B* **266**, 859 (1999).
- [19] T. House and M. J. Keeling, Insights from unifying modern approximations to infections on networks, *J. R. Soc. Interface* **8**, 67 (2011).
- [20] C.-R. Cai, Z.-X. Wu, M. Z. Chen, P. Holme, and J.-Y. Guan, Solving the dynamic correlation problem of the susceptible-infected-susceptible model on networks, *Phys. Rev. Lett.* **116**, 258301 (2016).
- [21] O. Ovaskainen, The quasistationary distribution of the stochastic logistic model, *J. Appl. Probab.* **38**, 898 (2001).
- [22] M. I. Dykman, I. B. Schwartz, and A. S. Landsman, Disease extinction in the presence of random vaccination, *Phys. Rev. Lett.* **101**, 078101 (2008).
- [23] O. Ovaskainen and B. Meerson, Stochastic models of population extinction, *Trends Ecol. Evol.* **25**, 643 (2010).
- [24] M. Assaf and B. Meerson, Extinction of metastable stochastic populations, *Phys. Rev. E* **81**, 021116 (2010).

- [25] J. Hindes and I. B. Schwartz, Epidemic extinction and control in heterogeneous networks, *Phys. Rev. Lett.* **117**, 028302 (2016).
- [26] M. Assaf and B. Meerson, WKB theory of large deviations in stochastic populations, *J. Phys. A: Math. Theor.* **50**, 263001 (2017).
- [27] J. Hindes and M. Assaf, Degree dispersion increases the rate of rare events in population networks, *Phys. Rev. Lett.* **123**, 068301 (2019).
- [28] J. Hindes, M. Assaf, and I. B. Schwartz, Outbreak size distribution in stochastic epidemic models, *Phys. Rev. Lett.* **128**, 078301 (2022).
- [29] J. Hindes, L. Mier-y Teran-Romero, I. B. Schwartz, and M. Assaf, Outbreak-size distributions under fluctuating rates, *Phys. Rev. Res.* **5**, 043264 (2023).
- [30] E. Korngut, O. Vilk, and M. Assaf, Weighted-ensemble network simulations of the susceptible-infected-susceptible model of epidemics, *Phys. Rev. E* **111**, 014146 (2025).
- [31] E. Korngut and M. Assaf, Impact of network assortativity on disease lifetime in the SIS model of epidemics, *Phys. Rev. E* **112**, 024302 (2025).
- [32] E. Volz, SIR dynamics in random networks with heterogeneous connectivity, *J. Math. Biol.* **56**, 293 (2008).
- [33] N. Sherborne, J. C. Miller, K. B. Blyuss, and I. Z. Kiss, Mean-field models for non-Markovian epidemics on networks, *J. Math. Biol.* **76**, 755 (2018).
- [34] N. T. Bailey, A statistical method of estimating the periods of incubation and infection of an infectious disease, *Nature (London)* **174**, 139 (1954).
- [35] K. Gough, The estimation of latent and infectious periods, *Biometrika* **64**, 559 (1977).
- [36] H. J. Wearing, P. Rohani, and M. J. Keeling, Appropriate models for the management of infectious diseases, *PLoS Med.* **2**, e174 (2005).
- [37] E. Cator, R. van de Bovenkamp, and P. van Mieghem, Susceptible-infected-susceptible epidemics on networks with general infection and cure times, *Phys. Rev. E* **87**, 062816 (2013).
- [38] P. van Mieghem and R. van de Bovenkamp, Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks, *Phys. Rev. Lett.* **110**, 108701 (2013).
- [39] M. Boguñá, L. F. Lafuerza, R. Toral, and M. Á. Serrano, Simulating non-Markovian stochastic processes, *Phys. Rev. E* **90**, 042108 (2014).
- [40] I. Z. Kiss, G. Röst, and Z. Vizi, Generalization of pairwise models to non-Markovian epidemics on networks, *Phys. Rev. Lett.* **115**, 078701 (2015).
- [41] G. Röst, Z. Vizi, and I. Z. Kiss, Impact of non-Markovian recovery on network epidemics, in *Biomat 2015* (World Scientific, Singapore, 2015), pp. 40–53.
- [42] I. Z. Kiss, J. C. Miller, and P. L. Simon, Non-Markovian epidemics, in *Mathematics of Epidemics on Networks: From Exact to Approximate Models* (Springer International Publishing, Cham, 2017), pp. 303–326.
- [43] G. Röst, Z. Vizi, and I. Kiss, Pairwise approximation for SIR-type network epidemics with non-Markovian recovery, *Proc. R. Soc. A* **474**, 20170695 (2018).
- [44] N. Masuda and L. E. Rocha, A Gillespie algorithm for non-Markovian stochastic processes, *SIAM Rev.* **60**, 95 (2018).
- [45] M. Starnini, J. P. Gleeson, and M. Boguñá, Equivalence between non-Markovian and Markovian dynamics in epidemic spreading processes, *Phys. Rev. Lett.* **118**, 128301 (2017).
- [46] M. Feng, S.-M. Cai, M. Tang, and Y.-C. Lai, Equivalence and its invalidation between non-Markovian and Markovian spreading dynamics on complex networks, *Nat. Commun.* **10**, 3748 (2019).
- [47] F. Ball, A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models, *Adv. Appl. Probab.* **18**, 289 (1986).
- [48] A. N. Startsev, On the distribution of the size of an epidemic in a non-Markovian model, *Theory Probab. Appl.* **41**, 730 (1997).
- [49] D. Clancy, SIR epidemic models with general infectious period distribution, *Stat. Probab. Lett.* **85**, 1 (2014).
- [50] R. R. Wilkinson and K. J. Sharkey, Impact of the infectious period on epidemics, *Phys. Rev. E* **97**, 052403 (2018).
- [51] T. Aquino and M. Dentz, Chemical continuous time random walks, *Phys. Rev. Lett.* **119**, 230601 (2017).
- [52] O. Vilk, R. Metzler, and M. Assaf, Non-Markovian gene expression, *Phys. Rev. Res.* **6**, L022026 (2024).
- [53] O. Vilk and M. Assaf, Escape from a metastable state in non-Markovian population dynamics, *Phys. Rev. E* **110**, 044132 (2024).
- [54] O. Vilk, M. Mobilia, and M. Assaf, Non-Markovian rock-paper-scissors games, *Phys. Rev. Res.* **7**, 023284 (2025).
- [55] M. I. Dykman, E. Mori, J. Ross, and P. Hunt, Large fluctuations and optimal paths in chemical kinetics, *J. Chem. Phys.* **100**, 5735 (1994).
- [56] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK, 2001).
- [57] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/bswp-wrfc> for more details on the model and additional results from the analysis and simulations.
- [58] This is justified since for $N \gg 1$, the epidemic duration is still much longer than the system's relaxation time [75].
- [59] Note that the leading-order WKB approximation here is valid as long as the total action, NS , is large [24,26,55].
- [60] In each realization, we sample the WT for both infection and recovery processes from the given WT distributions, execute the reaction that occurs first, advance the system time accordingly, update the distributions' mean rates, reset the timers (as done in Ref. [44] when the WT means are state dependent), and repeat until either disease extinction occurs or the maximum time is reached.
- [61] D. F. Anderson, A modified next reaction method for simulating chemical systems with time-dependent propensities and delays, *J. Chem. Phys.* **127**, 214107 (2007).
- [62] In these simulations, we performed a sufficiently large number of realizations so that, even after excluding those with a final outbreak fraction $< 10^{-2}$, we still retained the desired number of samples (see Figs. 1 and 2). This allowed us to probe the distribution around the mean without the influence of near-immediate extinction events.
- [63] Notably, the dependence of σ on x_i^* and α remains identical also when infection is exponential and recovery is gamma distributed, even though x_i^* changes in that case.
- [64] M. Assaf and M. Mobilia, Fixation of a deleterious allele under mutation pressure and finite selection intensity, *J. Theor. Biol.* **275**, 93 (2011).

- [65] J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, *et al.*, Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei Province, China: A descriptive and modelling study, *Lancet Infect. Dis.* **20**, 793 (2020).
- [66] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, *et al.*, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study, *Lancet Infect. Dis.* **20**, 911 (2020).
- [67] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data, *J. Clin. Med.* **9**, 538 (2020).
- [68] X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, Temporal dynamics in viral shedding and transmissibility of COVID-19, *Nat. Med.* **26**, 672 (2020).
- [69] S. T. Ali, L. Wang, E. H. Lau, X.-K. Xu, Z. Du, Y. Wu, G. M. Leung, and B. J. Cowling, Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions, *Science* **369**, 1106 (2020).
- [70] S. W. Park, B. M. Bolker, S. Funk, C. J. E. Metcalf, J. S. Weitz, B. T. Grenfell, and J. Dushoff, Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: Framework and applications to the novel coronavirus (SARS-CoV-2) outbreak, *J. R. Soc. Interface* **17**, 20200144 (2020).
- [71] H. Lee, G. Lee, T. Kim, S. Kim, H. Kim, and S. Lee, Variability in the serial interval of COVID-19 in South Korea: A comprehensive analysis of age and regional influences, *Front. Public Health* **12**, 1362909 (2024).
- [72] F. Carrat, E. Vergu, N. M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A.-J. Valleron, Time lines of infection and disease in human influenza: A review of volunteer challenge studies, *Am. J. Epidemiol.* **167**, 775 (2008).
- [73] A. W. Byrne, D. McEvoy, A. B. Collins, K. Hunt, M. Casey, A. Barber, F. Butler, J. Griffin, E. A. Lane, C. McAloon, *et al.*, Inferred duration of infectious period of SARS-CoV-2: Rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases, *BMJ Open* **10**, e039856 (2020).
- [74] I. Voinsky, G. Baristaite, and D. Gurwitz, Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients, *J. Infect.* **81**, e102 (2020).
- [75] M. Turkyilmazoglu, Explicit formulae for the peak time of an epidemic from the SIR model, *Physica D* **422**, 132902 (2021).