Fock-space approach to stochastic susceptible-infected-recovered models

Danillo B. de Souza,[1] Hugo A. Araújo [2,3] Gerson C. Duarte-Filho [4] Eamonn A. Gaffney,[5]
Fernando A. N. Santos,[2,6] and Ernesto P. Raposo[3]

[1]*Basque Center for Applied Mathematics, Mathematical, Computational and Experimental Neuroscience Research Group,*
48009 Bilbao, Bizkaia, Basque-Country, Spain
[2]*Departamento de Matemática, Universidade Federal de Pernambuco, 50670-901 Recife, PE, Brazil*
[3]*Laboratório de Física Teórica e Computacional, Departamento de Física, Universidade Federal de Pernambuco,*
50670-901 Recife, PE, Brazil
[4]*Departamento de Física, Universidade Federal de Sergipe, 49100-000 São Cristóvão, SE, Brazil*
[5]*Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom*
[6]*Department of Anatomy and Neurosciences, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Neuroscience,*
1081 HZ Amsterdam, The Netherlands

We investigate the stochastic susceptible-infected-recovered (SIR) model of infectious disease dynamics in the Fock-space approach. In contrast to conventional SIR models based on ordinary differential equations for the subpopulation sizes of S, I, and R individuals, the stochastic SIR model is driven by a master equation governing the transition probabilities among the system's states defined by SIR occupation numbers. In the Fock-space approach the master equation is recast in the form of a real-valued Schrödinger-type equation with a second quantization Hamiltonian-like operator describing the infection and recovery processes. We find exact analytic expressions for the Hamiltonian eigenvalues for any population size N. We present small- and large-N results for the average numbers of SIR individuals and basic reproduction number. For small N we also obtain the probability distributions of SIR states, epidemic sizes and durations, which cannot be found from deterministic SIR models. Our Fock-space approach to stochastic SIR models introduces a powerful set of tools to calculate central quantities of epidemic processes, especially for relatively small populations where statistical fluctuations not captured by conventional deterministic SIR models play a crucial role.

I. INTRODUCTION

The dynamics of infectious diseases in a population has been long addressed successfully through susceptible-infected-recovered (SIR) and related models [1–10], since the seminal work by Kermack and McKendrick in 1927 [1]. SIR models belong to a class of compartmental models in which the population is generally subdivided into three categories of individuals: those who are susceptible to catch the disease (S), those who are infected and can spread the disease to susceptible individuals (I), and those who have become immune (recovered) or died (removed) and can no longer spread or catch the disease (R).

From a historical perspective, Kermack and McKendrick [1] introduced the first version of a SIR model based on a set of coupled ordinary differential equations that drive the dynamics of the subpopulations of susceptible, infected, and recovered or dead individuals in an epidemic. In this early model it was assumed at the individual level that at any given time each individual is in one of the three possible states (S, I, or R). In a deterministic mean-field-type formulation at the population level, a system of coupled differential equations on the average subpopulations sizes arises. In addition, the population size is fixed (no births or deaths by causes other than the disease itself), the infectious agent has no incubation time, the duration of infectiousness matches the duration of being

infected, and the population is homogeneous, with no age, spatial, or social structure. The authors then compared [1] with relative success the model solution with data on the number of deaths with time during the Bombay plague of 1905–1906.

Over essentially the past 100 years the original SIR model has been substantially improved to address the time evolution of a significant variety of epidemic processes [2–45]. For example, rather than being governed by a set of ordinary differential equations with each individual in a given state at any time as in Ref. [1], the underlying dynamics of *stochastic* SIR models is driven by a master equation for the probability distribution of the system's states, which evolve in time according to a set of transition rates among them. In this context, stochastic SIR models are necessary to describe the nondeterministic variability of the epidemic size and duration, together with the probability of disease spread, particularly in smaller communities. Such features have become highly relevant in current times, especially concerning the epidemic control and role of vaccination and prophylaxis campaigns [42–45].

The relevant capabilities of resolving epidemic fluctuation dynamics for smaller populations through stochastic SIR models, however, requires extensive technical and computational effort compared to the model dynamics driven by ordinary differential equations of deterministic SIR models. In particular, while Schutz *et al.*'s solution [46] for SIR

interactions on a chain with homogeneous initial conditions is a notable exception, typically analytical solutions for any fluctuation-based phenomena associated with SIR systems are absent and more complex numerical approaches are required. For the coarser-grained stochastic differential equation framework, computations typically use Euler-Maruyama [35,47], implicit Euler [38,42], or Milstein [47,48] methods, while for models capable of resolving at the level of the individual one finds numerical techniques based on Gillespie's algorithm [49,50] or direct Monte Carlo simulation [18,22] (see also Ref. [51]). However, these computational frameworks necessitated by the modeling at the resolution of the individual are not generally amenable to likelihood maximization or Bayesian inference. We notice that this is particularly problematic on confronting such models with data as parameter inference, hypothesis testing and model selection then require computationally very demanding, likelihood-free methods, such as approximate Bayesian computation [52].

Hence, our objective is to apply the so-called Fock-space approach [53,54] to the study of individual-based stochastic SIR models without the need of stochastic simulation or Monte Carlo algorithms that prevent the general use of likelihood methods when comparing modeling with data. In particular, this entails that we do not seek to simulate or solve the associated stochastic differential or master equations but instead to utilize the Fock-space method of quantum physics, which relies on tools inherited from the second quantization formalism combined with symbolic algebra. Within this framework, the Fock-space solution for a population of size N in time t is obtained in terms of the eigenvalues and eigenvectors of a Hamiltonian-like operator that drives the stochastic SIR model dynamics. Here we find exact analytical expressions for the eigenvalues for any N. The Hamiltonian eigenvectors are obtained for values of N that are not excessively large using a symbolic computing software. From the sets of eigenvalues and eigenvectors the dynamics of all important quantities of interest is determined, such as time-dependent probabilities and average values of the number of susceptible, infected, and recovered individuals, mean size and duration of the epidemic, and basic reproduction number. In particular, for small N closed-form analytical expressions for these quantities are feasible.

The Fock-space formalism was first proposed [55,56] by Schönberg in 1952 (see also Ref. [57]) and later rediscovered [53,54] by Doi in the context of diffusion-controlled processes in liquid media and chemical reactions. In this approach the master equation of a set of general random particles is written in the form of a real-valued Schrödinger-type equation, with the probability to find the system in a given state at a certain time playing a role similar to that of the wave function in quantum mechanics. The Hamiltonian-like operator in the Fock-space approach is written in a basis of discrete Fock states associated with the occupation numbers of the system's constituents (for example, the numbers of each type of molecule in a chemical reaction or particles at each site of a discrete lattice). The Fock-space formalism was later successfully extended to treat a variety of other stochastic systems, for instance, gene expression [58], absorbing states in nonequilibrium lattice dynamics [59], general reaction-diffusion dynamics [60], and spins chains [61]. More recently, our group has applied the Fock-space approach to study chem-

ical enzymes interactions [62], fermionic diffusion [63], and the random search problem [64,65].

A general limitation of the Fock-space method concerns the diagonalization via symbolic computation of large Hamiltonian-like matrices for high numbers of constituents. Nevertheless, even in this case a number of results can still be obtained without the need of diagonalization of the Hamiltonian, as shown in this work. Further, as discussed above, we also remark that the gain in addressing the epidemic dynamics problem through stochastic SIR models is more significant precisely with small populations, in which statistical fluctuations of central quantities—not captured by deterministic SIR models—play a crucial role.

We notice that recent studies [36,37] have applied second quantization ideas to address disease dynamics in a population. However, while in Ref. [36] a simpler model (SI, without recovered individuals) is investigated, with focus on the average sizes of the susceptible and infected subpopulations through diagrammatic expansion in small networks, in Ref. [37] the second quantization approach is specifically applied with the aim to find the set of ordinary differential equations for the average subpopulations in a SIR dynamics. In contrast, in our work these average values can be also obtained from the eigenvalues and eigenvectors of the Hamiltonian-like operator combined with the time evolution of the system's state vector. Indeed, here we calculate, for example, the distributions of subpopulations sizes, epidemic sizes, and durations for populations that are not excessively large, noting these cannot be found by simply solving the ordinary differential equations as in Ref. [37].

This article is organized as follows. In Sec. II we review the general formalism of the Fock-space approach. The method is applied to the stochastic SIR model in Sec. III. General expressions for some relevant quantities are provided, including the exact Hamiltonian eigenvalues as functions of the model parameters (infection and recovery rates) and population size N. Fock-space results for $N = 20$, $N = 35$, and $N = 10^4$ are discussed in Sec. IV. Last, final remarks and conclusions are left to Sec. V. We also include Appendices A and B with details on a novel derivation of the basic reproduction number for this model and an illustrative example of the calculation of some closed-form analytical results for $N = 3$, respectively.

II. THE FOCK-SPACE APPROACH

We begin by briefly reviewing the general formalism of the Fock-space approach [53,54]. Consider a system with N constituents (e.g., molecules undergoing a chemical reaction or individuals in a population) that can be grouped into k subsets of distinct species. We denote by $N_j(t)$ the number of constituents of species j at time t, with $j = 1, 2, \ldots, k$, and $\sum_{j=1}^{k} N_j(t) = N$.

The species are allowed to interact through one or more processes labeled i that occur at given rates r_i. For deterministic continuum models, one has fractions m_{ij} of the constituents of each species j that generate, via a law of mass action for every process i, a new set of fractions n_{ij}, as generally described by

$$\sum_{j=1}^{k} m_{ij} N_j(t) \xrightarrow{r_i} \sum_{j=1}^{k} n_{ij} N_j(t). \tag{1}$$

In turn, these interactions give rise to dynamical systems of ordinary differential equations, yielding the continuum model.

In a stochastic dynamic evolution, one typically considers the probability $\mathcal{P}(\mathbf{N}, t)$ to find the system at the state $\mathbf{N}(t) = (N_1(t), \ldots, N_k(t))$ in time t. The associated master equation reads

$$\frac{\partial \mathcal{P}(\mathbf{N}, t)}{\partial t} = \sum_{\mathbf{N}'} [\tau_{\mathbf{N}' \to \mathbf{N}} \mathcal{P}(\mathbf{N}', t) - \tau_{\mathbf{N} \to \mathbf{N}'} \mathcal{P}(\mathbf{N}, t)], \quad (2)$$

where in the case of a Markovian process, as assumed in this work, the transition rates $\tau_{\mathbf{N}' \to \mathbf{N}}$ from state $\mathbf{N}'$ to $\mathbf{N}$ are time independent.

To start exploring the Fock-space tools, we consider that a state of the system can be represented in a Fock space $\mathcal{F}$ obtained by the direct product of the Hilbert spaces $\mathcal{S}_j$ of all species j. That is, $\mathcal{F} = \mathcal{S}_1 \otimes \ldots \otimes \mathcal{S}_k$, with $\mathcal{S}_j = \{1, \ldots, N\}$. In Dirac notation, $|n\rangle = |s_1 \ldots s_k\rangle$ represents a pure Fock state, i.e., a Fock state with well-defined occupation numbers $s_1 \in S_1$ of constituents of species $j = 1$, $s_2 \in S_2$ of species $j = 2$, and so on. The label n indexes the states in some given order and the set $\{|n\rangle\}$ of all pure Fock states provides a basis for the Fock space $\mathcal{F}$.

By following Refs. [53,54] and rewriting $\mathcal{P}(\mathbf{N}, t)$ in the new notation $P(n, t)$, with $n = \{s_1 \ldots s_k\}$, the statistical description of the stochastic system in time t can be characterized by the state vector

$$|\Psi(t)\rangle = \sum_n P(n, t) |n\rangle, \quad (3)$$

which comprises a linear superposition with each pure Fock state weighted by the respective time-dependent probability, so that $\sum_n P(n, t) = 1$ for any t.

The creation ($\alpha_j^\dagger$) and annihilation (α_j) operators for each species j act on the pure Fock states respectively according to [53,54]

$$\begin{aligned} \alpha_j^\dagger |n\rangle &= |s_1 \ldots (s_j + 1) \ldots s_k\rangle, \\ \alpha_j |n\rangle &= s_j |s_1 \ldots (s_j - 1) \ldots s_k\rangle, \end{aligned} \quad (4)$$

with $\alpha_j^\dagger \alpha_j$ identifying the number operator of the constituents of species j. If we denote the vacuum state (absence of constituents of any species) by $|0\rangle$, then $\alpha_j |0\rangle = 0$ and $\langle 0| \alpha_j^\dagger = 0$. It is thus straightforward to verify the commutation rules $[\alpha_i, \alpha_j^\dagger] = \delta_{ij}$, $[\alpha_i, \alpha_j] = 0$, and $[\alpha_i^\dagger, \alpha_j^\dagger] = 0$, where δ stands for the Kronecker delta. The combination of these results leads to the orthogonality property of the pure Fock states.

By considering Eq. (3), the master equation (2) can be recast in the form of a real-valued ($i\hbar \equiv 1$) Schrödinger-type equation,

$$\frac{\partial |\Psi(t)\rangle}{\partial t} = -H(\alpha_1^\dagger, \alpha_1, \ldots, \alpha_k^\dagger, \alpha_k) |\Psi(t)\rangle, \quad (5)$$

with the Hamiltonian-like operator H written as a function of the set $\{\alpha_j^\dagger, \alpha_j\}$ and dependent on the transition rates to be consistent with Eq. (2). From Eq. (5) the state vector dynamics is given by

$$|\Psi(t)\rangle = U(t) |\Psi(0)\rangle, \quad (6)$$

where the time evolution operator is

$$U(t) = \exp(-H(\alpha_1^\dagger, \alpha_1, \ldots, \alpha_k^\dagger, \alpha_k)t), \quad (7)$$

and $|\Psi(0)\rangle$ is the system's initial state vector, see Eq. (3). In the stochastic SIR model (see below), we remark that H in Eq. (5) is an infinitesimal stochastic operator of non-Hermitian type with null column sum on the Fock-space basis, whereas U in Eq. (7) is a nonunitary stochastic operator with unit column sum. For further details concerning the use of operators in stochastic dynamics, see Refs. [66–70] and references therein.

After writing H explicitly in terms of the creation and annihilation operators, and expressing it in a matrix form on the basis $\{|n\rangle\}$ of pure Fock states, the set of tools from quantum mechanics can be employed in the Fock-space representation to provide the time evolution and average values of all relevant observables of the stochastic system, as described in the following.

III. FOCK-SPACE APPROACH APPLIED TO THE STOCHASTIC SIR MODEL

We now apply the Fock-space formalism to the stochastic SIR model. As mentioned, the SIR model concerns an epidemic taking place in a population of N individuals that can be subdivided into $k = 3$ distinct groups: susceptible (S), infected (I), and recovered (R). In analogy to Eq. (1), the subpopulations interact through the processes

$$\begin{aligned} S + I &\xrightarrow{\alpha} 2I, \\ I &\xrightarrow{\beta} R, \end{aligned} \quad (8)$$

where the two model parameters $\alpha > 0$ and $\beta > 0$ (in units of t^{-1}) represent, respectively, the infection rate at which susceptible individuals become infected by the contact with a previously infected one, and the recovery rate at which infected individuals become recovered (immune or deceased). In general, the larger β is in comparison to α, the shorter the epidemic lasts on average.

In the stochastic SIR model the sizes of the three subpopulations are determined statistically from a t-dependent probability distribution driven by a master equation in the form of Eq. (2). Therefore, the stochastic SIR model with infection and recovery processes as in (8) can be properly addressed in a Fock-space approach.

We depict the pure Fock states of the stochastic SIR model in Dirac notation by $|n\rangle = |s \, i \, r\rangle$, with

$$s, i, r \in \{0, 1, \ldots, N\} \quad (9)$$

as the respective numbers of susceptible, infected, and recovered individuals. A basis of the Fock space,

$$\mathcal{F} = \mathcal{S} \bigotimes \mathcal{I} \bigotimes \mathcal{R}, \quad (10)$$

can be thus built by taking into account the set $\{|s \, i \, r\rangle\}$ of all kets constrained to the total population size, $s + i + r = N$, generating $N_h = (N + 1)(N + 2)/2$ distinct possibilities.

Following Eq. (3), a general state vector of the system is written as

$$|\Psi(t)\rangle = \sum_{s,i,r} P(s, i, r, t) |s \, i \, r\rangle, \quad (11)$$

with the constraint $s + i + r = N$ implied in the summations above, and

$$P(s, i, r, t) = \langle s\, i\, r | \Psi(t) \rangle \tag{12}$$

denoting the probability of the system occupying the state with s susceptible, i infected, and r recovered individuals in time t, so that $\sum_{s,i,r} P(s, i, r, t) = 1$ for any t. Further, from Eq. (4) the creation ($s^\dagger$) and annihilation (s) operators associated with the number of susceptible individuals act on the pure Fock states respectively via

$$\begin{aligned} s^\dagger |n\rangle &= |(s+1)\, i\, r\rangle, \\ s |n\rangle &= s |(s-1)\, i\, r\rangle, \end{aligned} \tag{13}$$

with similar expressions for the operators related to the infected ($i^\dagger$, i) and recovered ($r^\dagger$, r) subpopulations.

The consistency between the master and Schrödinger-type equations of the stochastic SIR model driven by processes (8) yields the Hamiltonian-like operator

$$H = -\alpha[(i^\dagger)^2 - s^\dagger i^\dagger]si - \beta(r^\dagger - i^\dagger)i, \tag{14}$$

with matrix elements on the basis $\{|n\rangle = |s\, i\, r\rangle\}$ of pure Fock states, $h_{mn} = \langle m| H |n\rangle$, given by

$$\begin{aligned} h_{mn} = i_n \big[&\alpha\, s_n \big(\delta_{i_m, i_n} \delta_{r_m, r_n} \delta_{s_m, s_n} - \delta_{r_m, r_n} \delta_{i_m, i_n+1} \delta_{s_m, s_n-1} \big) \\ &+ \beta\, \delta_{s_m, s_n} \big(\delta_{i_m, i_n} \delta_{r_m, r_n} - \delta_{i_m, i_n-1} \delta_{r_m, r_n+1} \big) \big]. \end{aligned} \tag{15}$$

Above the notation s_n means that the state $|n\rangle$ comprises s_n susceptible individuals and so forth. Since N_h is the number of pure Fock states, we note that the matrix h has dimension $N_h \times N_h$.

We next diagonalize the Hamiltonian-like matrix h to obtain the sets of eigenvalues $\{\lambda_\nu\}$ and right eigenvectors $\{|\lambda_\nu\rangle\}$ of the stochastic SIR model, with $\nu = 1, 2, \ldots, N_h$. The eigenvalues are given by the roots of the characteristic polynomial $p(\lambda)$ of matrix h, which can be written as

$$p(\lambda) = \lambda^{N+1} \prod_{r=1}^{N} p_r(\lambda), \tag{16}$$

with N auxiliary polynomials $p_r(\lambda)$ of degree r, thus implying that $p(\lambda)$ is actually degree N_h, as expected.

Some of the main results of this work regard the exact analytical expressions for the polynomials and eigenvalues. Indeed, the structure of the matrix elements (15) of the stochastic SIR model allows us to explicitly write the auxiliary polynomials as

$$\begin{aligned} p_{r=1}(\lambda) &= \lambda - \beta, \\ p_{r=2}(\lambda) &= [\lambda - (\alpha + \beta)](\lambda - 2\beta), \\ p_r(\lambda) &= \prod_{k=0}^{r-1} (\lambda - \lambda_{rk}), \quad r = 3, \ldots, N, \quad k = 0, \ldots, r-1, \end{aligned} \tag{17}$$

with the eigenvalues expressed in exact form by

$$\begin{aligned} \lambda_{r=1} &= \beta, \\ \lambda_{r=2, k=0} &= \alpha + \beta, \quad \lambda_{r=2, k=1} = 2\beta, \\ \lambda_{rk} &= a_{rk}\alpha + b_{rk}\beta, \quad r = 3, \ldots, N, \quad k = 0, \ldots, r-1, \end{aligned} \tag{18}$$

and the factors

$$\begin{aligned} a_{rk} &= -k^2 + (r-2)k + r - 1, \\ b_{rk} &= k + 1. \end{aligned} \tag{19}$$

We therefore remark that the set of eigenvalues $\{\lambda_\nu\}$ of the stochastic SIR model in the Fock-space approach can be readily expressed exactly as functions of the parameters α and β by Eqs. (18) and (19), for any population size N.

For practical purposes here we choose the ordering $\nu = 1, 2, \ldots, N_h$ of the eigenvalues $\{\lambda_\nu\}$ from the smallest to the largest one (see Appendix B for an illustrative example with $N = 3$ and $N_h = 10$). One important feature from Eqs. (18) and (19) is that all eigenvalues are either null or positive. In fact, the pure Fock states $|N - j, 0, j\rangle$, with $i = 0$ infected individuals and $j = 0, 1, \ldots, N$, are eigenvectors of h with null eigenvalue, since the dynamics (8) ceases when the disease transmission is no longer possible due to the absence of infectious cases. We thus write $\lambda_\nu = 0$ for $\nu = j + 1 = 1, 2, \ldots, N + 1$, and $\lambda_\nu > 0$ for $\nu = N + 2, \ldots, N_h$. The remaining eigenvectors $\{|\lambda_\nu\rangle\}$, $\nu = N + 2, \ldots, N_h$, can be in principle determined on the basis $\{|s\, i\, r\rangle\}$ as functions of α and β by using a symbolic computing software (in this work we have used *Mathematica*).

The epidemic is generally considered to start with $i = i_0$ infected, $r = r_0 = 0$ recovered, and $s = s_0 = N - i_0$ susceptible individuals in $t = 0$, i.e., the initial state vector is usually $|\Psi(0)\rangle = |s_0\, i_0\, r_0\rangle = |N - i_0, i_0, 0\rangle$. We can write the initial state vector as well in the form of a linear superposition of right eigenvectors,

$$|\Psi(0)\rangle = \sum_\nu a_\nu |\lambda_\nu\rangle, \tag{20}$$

with coefficients a_ν, and by combining Eqs. (6), (7), and (20) the system's state vector in time t is expressed as

$$|\Psi(t)\rangle = \sum_\nu a_\nu e^{-\lambda_\nu t} |\lambda_\nu\rangle, \tag{21}$$

from which all significant quantities associated with the stochastic SIR model can be determined.

Indeed, by considering Eqs. (12) and (21) all statistical moments of the distributions of numbers of susceptible, infected, and recovered individuals can be obtained. For example, by denoting as $\langle S \rangle$ the average number of susceptible individuals we find

$$\langle S \rangle(t) = \sum_{s,i,r} s\, P(s, i, r, t), \tag{22}$$

with the constrained sums above and analogous expressions for the average numbers of infected $[\langle I \rangle(t)]$ and recovered $[\langle R \rangle(t)]$ individuals, implying $\langle S \rangle(t) + \langle I \rangle(t) + \langle R \rangle(t) = N$ at any time t.

Three other relevant quantities are the total size and duration of an epidemic, and the basic reproduction number.

Let us denote by $p(\eta)$ the probability that an epidemic has total size η. One possible way to calculate $p(\eta)$ is [2] to consider the number r of recovered individuals in the steady-state regime, so that all infected cases that emerged throughout the epidemic have had enough time to recover, that is, $\eta = r$ and $i = 0$ as $t \to \infty$. If we opt not to count in η the i_0 individuals

already infected in $t = 0$, then

$$p(\eta) = \lim_{t \to \infty} P(N - i_0 - \eta, \, 0, \, i_0 + \eta, \, t). \qquad (23)$$

By projecting the state vector $|\Psi(t)\rangle$, Eq. (21), in the steady-state regime onto the pure Fock state with $s = N - i_0 - \eta$, $i = 0$, and $r = i_0 + \eta$, we find from Eq. (12) the probability distribution of an epidemic of total size η,

$$p(\eta) = \sum_\nu a_\nu \, \langle N - i_0 - \eta, \, 0, \, i_0 + \eta | \lambda_\nu \rangle, \qquad (24)$$

with the sum only over null eigenvalues. Since in our ordering the null eigenvalues $\lambda_\nu = 0$ are associated with the eigenvectors $|N - \nu + 1, \, 0, \, \nu - 1\rangle$, for $\nu = 1, 2, \ldots, N + 1$, then $p(\eta) = a_{\nu = \eta + i_0 + 1}$. From this result the epidemic's average size $\langle \eta \rangle$ is readily obtained as a function of the parameters α and β of the stochastic SIR model.

The Fock-space approach also provides a way to estimate the mean duration of an epidemic. Consider the subset $\{|N - j, \, 0, \, j\rangle\}$ of pure Fock states (also eigenvectors) with $i = 0$ infected individuals, where $j = 0, 1, \ldots, N$. The probability that none of these states has been reached in time $t = \tau$, so that the epidemic have persisted up to this time, is given by

$$p_{\text{not}}(\tau) = 1 - \sum_j \langle N - j, \, 0, \, j | \Psi(\tau) \rangle. \qquad (25)$$

Therefore, the probability that the epidemic ceases at some subsequent time is $1 - p_{\text{not}}(\tau)$, and so the probability density that the epidemic duration equals τ reads

$$\rho(\tau) = \frac{\partial}{\partial \tau} \sum_j \langle N - j, \, 0, \, j | \Psi(\tau) \rangle. \qquad (26)$$

Again, all statistical moments of $\rho(\tau)$ of the stochastic SIR model can be determined from Eqs. (21) and (26). In particular, the average duration of the epidemic is calculated in the Fock-space approach as

$$\langle \tau \rangle = \int_0^\infty \tau \rho(\tau) d\tau = - \sum_{j, \nu} \frac{a_\nu}{\lambda_\nu} \langle N - j, \, 0, \, j | \lambda_\nu \rangle, \qquad (27)$$

with the sum in ν extending over non-null eigenvalues λ_ν.

We end this section by considering the basic reproduction number $\mathcal{R}_0$ of an epidemic process, generally defined [2] as the expected number of infected cases caused by a single infected individual in a completely susceptible population. So $\mathcal{R}_0$ gives a measure of the potential for disease spread in a population: the larger the value of $\mathcal{R}_0$, the easier the epidemic spreads and the harder its control becomes.

Complying with usual notation [2], we denote $b(a)$ to be the average number of individuals that caught the disease from a single individual (patient zero), who remained infectious from $t = 0$ to $t = a$. In addition we further define $F(a)$ to be the probability that a newly infected individual has remained infectious during this time interval. In the asymptotic steady-state limit $t \to \infty$ one thus has [2]

$$\mathcal{R}_0 = \int_0^\infty b(a) F(a) da. \qquad (28)$$

In the case of the stochastic SIR model with infection (α) and recovery (β) rates defined in processes (8), we note that [35]
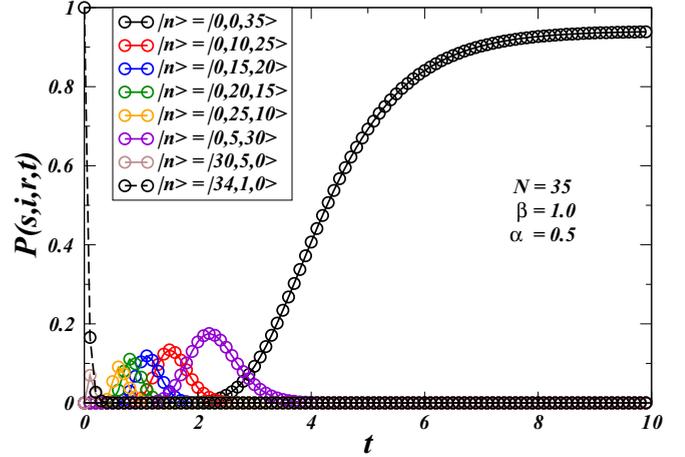
FIG. 1. Probability $P(s, i, r, t)$ of some selected Fock states $|n\rangle = |s, i, r\rangle$ with s susceptible, i infected, and r recovered individuals in a population of size $N = 35$ as a function of time t, for infection and recovery rates $\alpha = 0.5$ and $\beta = 1.0$, respectively. One individual was initially infected, $i_0 = 1$, as indicated by the maximum in $P(34, 1, 0, t)$ in $t = 0$ (dashed line with black circles). $P(0, 0, 35, t)$ (solid line with black circles) approaches saturation with all previously infected individuals recovered in the $t \to \infty$ steady-state limit. The solid and dashed lines are to guide the eye in interpreting the overall behavior.

$b(a) = \alpha \langle S \rangle(a)$ and $F(a) = e^{-\beta a}$, implying

$$\mathcal{R}_0 = \alpha \int_0^\infty e^{-\beta a} \langle S \rangle(a) da, \qquad (29)$$

with $\langle S \rangle(a)$ given in the Fock-space approach by Eq. (22). It is also possible to convert the integral above into a Riemann sum over discrete unit time intervals to speed up the calculations with negligible difference in the steady-state limit.

In Appendix A we present a novel derivation of Eq. (29). We stress that this is an original contribution for the calculation of the basic reproduction number $\mathcal{R}_0$ in the context of the Fock-space approach to the stochastic SIR model.

IV. RESULTS AND DISCUSSION

We now present results of the quantities worked in the previous section for populations of small ($N = 20$ and $N = 35$) and large ($N = 10^4$) sizes.

A. Cases $N = 20$ and $N = 35$

We start by considering the case of a population with $N = 35$ susceptible individuals and a single one infected at the beginning of the epidemic, $i_0 = 1$ in $t = 0$. This means that the initial state of the system is $|\Psi(0)\rangle = |s_0 i_0 r_0\rangle = |34, 1, 0\rangle$.

In Fig. 1 and Fig. 2 we have chosen the infection and recovery rates as $\alpha = 0.5$ and $\beta = 1.0$, respectively. From Eq. (8) this corresponds to the situation in which the epidemic should evolve toward a regime with most individuals recovered on average at sufficiently long times.

By calculating the eigenvalues $\{\lambda_\nu\}$ and right eigenvectors $\{|\lambda_\nu\rangle\}$ following the prescription of last section, the system dynamics is driven by Eqs. (20) and (21). Figure 1 presents in
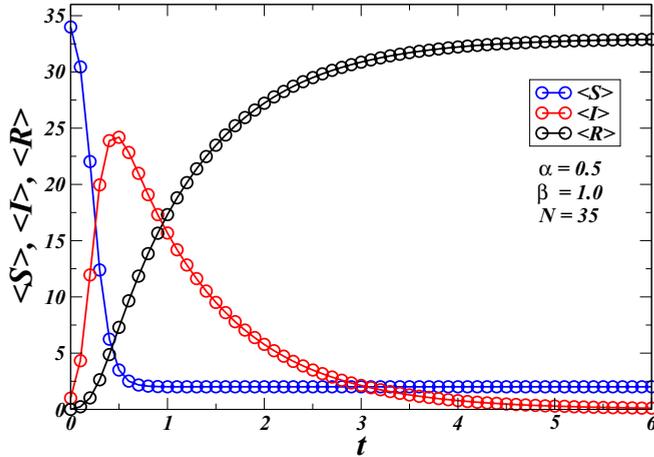
FIG. 2. Average numbers of susceptible ($\langle S \rangle$), infected ($\langle I \rangle$), and recovered ($\langle R \rangle$) individuals in a population of size $N = 35$ as a function of time t, for infection and recovery rates $\alpha = 0.5$ and $\beta = 1.0$, respectively. The mean number of infected cases (red) peaks at an intermediate time and then decreases to zero as the $t \to \infty$ saturation limit approaches with all previously infected individuals recovered. The solid lines are to guide the eye in interpreting the overall behavior.

FIG. 3. Probability $p(\eta)$ of total epidemic size η in a population with $N = 35$ individuals and $i_0 = 1$ initially infected case, for various recovery rates β and fixed infection rate $\alpha = 1$. For $\beta \gtrsim 25$ very few individuals become infected as $p(\eta) \approx 0$ for $\eta \gtrsim 1$. In contrast, for $\beta \ll 10$ the entire population tends to be infected at some point of the epidemic, since the height of the maximum at $\eta = N - i_0 = 34$ increases for lower β in this regime. For intermediate β this maximum occurs at $0 < \eta < N - i_0$, indicating that the epidemic does not reach all individuals but a fraction of the population. The solid lines are to guide the eye in interpreting the overall behavior.

circles some probability distributions $P(s, i, r, t)$, Eq. (12), as a function of time t (with the solid and dashed lines to guide the eye in interpreting the overall behavior). While the probability $P(34, 1, 0, t)$ of the initial state with $i_0 = 1$ infected individual rapidly decreases from one to zero (dashed line with black circles), we note that the probability $P(0, 0, 35, t)$ to have the full population recovered ($r = N = 35$) grows progressively with time (solid line with black circles). We also observe in Fig. 1 that the probability of pure Fock states with intermediate numbers of recovered individuals (colored symbols) displays a maximum at values of t that increase with r.

The average numbers of susceptible ($\langle S \rangle$), infected ($\langle I \rangle$), and recovered ($\langle R \rangle$) individuals are calculated from Eq. (22) and analogous equations, and plotted in circles as a function of t in Fig. 2. We notice at any time that $\langle S \rangle(t) + \langle I \rangle(t) + \langle R \rangle(t) = N$, as expected. We also observe that the initial growth of $\langle I \rangle$ eventually recedes, giving way to an increase in $\langle R \rangle$ toward the limit of full recovery of previously infected cases, consistent with Fig. 1.

In Fig. 3 we investigate the variability of the total size η of the epidemic (not counting the initially infected individual, $i_0 = 1$) by plotting the corresponding probability distribution $p(\eta)$, Eq. (24). We now keep $\alpha = 1$ fixed and show curves of $p(\eta)$ versus η for various β. Interestingly, we observe a dominant maximum at $\eta = 0$ for high recovery rates $\beta \gtrsim 25$, indicating that in this regime very few individuals are infected, yielding $p(\eta) \approx 0$ for $\eta \gtrsim 1$. On the other hand, for $\beta \ll 10$ the entire population tends to be infected by the disease and recovered at some point of the epidemic dynamics (up to the steady-state limit), as shown by the increase of the height of the second maximum at $\eta = N - i_0 = 34$ for lower β in this regime. For intermediate β this second maximum takes place at $0 < \eta < N - i_0$, showing that in this case it is more

likely that the epidemic does not reach all individuals, but a somewhat considerable fraction of the population.

The average duration of the epidemic $\langle \tau \rangle$ and basic reproduction number $\mathcal{R}_0$ as a function of α are shown, respectively, in Fig. 4 and Fig. 5 for $\beta = 0.5, 1.0, 2.0$. To proceed with less time consuming computation we considered the Riemann
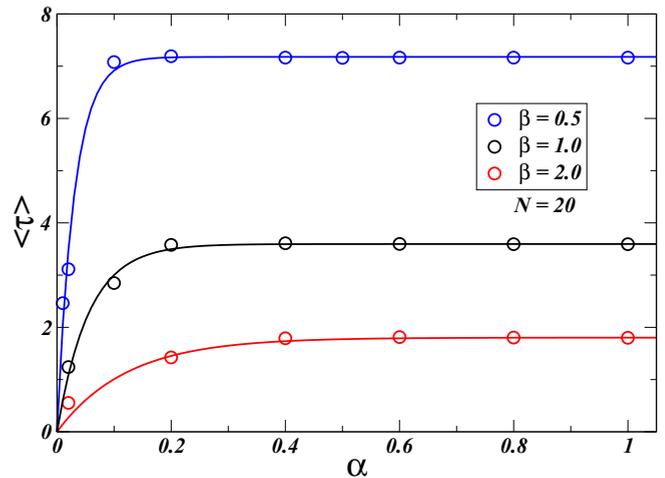
FIG. 4. Mean epidemic duration $\langle \tau \rangle$ as a function of the infection rate α for some fixed recovery rates β, in a population of size $N = 20$ with $i_0 = 1$ individual initially infected. $\langle \tau \rangle$ grows with α but rapidly saturates, indicating that a further increase of α in the strongly infectious regime does not impact significantly the epidemic dynamics of a small population. The solid lines are to guide the eye in interpreting the overall behavior.
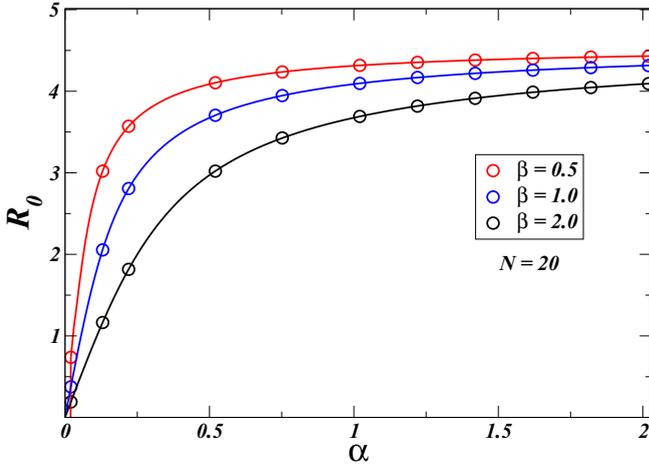
FIG. 5. Basic reproduction number $\mathcal{R}_0$ as a function of the infection rate α for some fixed recovery rates β, in a population of size $N = 20$ with $i_0 = 1$ individual initially infected. Likewise Fig. 4, $\mathcal{R}_0$ also increases with α, though with a somewhat lower growth rate, as epidemics are harder to control in stronger infectious regimes. The solid lines are to guide the eye in interpreting the overall behavior.

sum form of the integrals in Eqs. (27) and (29) with $N = 20$ individuals.

We observe in Fig. 4 that the mean epidemic duration increases with α for fixed β, but nearly saturates for $\alpha \gtrsim \beta$. Likewise, the basic reproduction number also grows with α, see Fig. 5, indicating that the trend for disease spread is higher for larger α, as expected. In both cases we note that increasing α for fixed β in the strongly infectious regime does not make a significant impact on the mean duration and basic reproduction number of the epidemic process.

B. Case $N = 10^4$

Finally, we turn to the study of the epidemic evolution in a large population with $N = 10^4$ individuals. We aim here at building the large-N correspondence between the stochastic SIR model in the Fock-space approach and the deterministic SIR model. We also make explicit below the influence of fluctuations by comparing analytical expressions for the basic reproduction number of these models for not so large populations.

As discussed, for large N relative fluctuations of order $O(1/\sqrt{N})$ around the mean value of key quantities are not as relevant when compared to the previous cases of small populations. So the gain in addressing the epidemic dynamics problem in the large-N regime through stochastic SIR models is not so significant if contrasted with the approach from deterministic SIR models. Furthermore, when applying the Fock-space method with large populations, the dimension $N_h \times N_h$ of the Hamiltonian-like matrix h grows as $N_h \approx N^2$, hampering considerably the symbolic computation of the eigenvectors. Nevertheless, by combining [37] the second quantization formalism with the statistical properties [2] of the moment generating function associated with $P(s, i, r, t)$, a number of results in the large-N regime can still be obtained

in the Fock-space approach as follows, without the need to diagonalize h.

The generating function of the moments of the distribution $P(s, i, r, t)$ is defined [2] as

$$M(\theta_s, \theta_i, \theta_r, t) = \sum_{s,i,r} P(s, i, r, t) e^{\theta_s s + \theta_i i + \theta_r r}, \quad (30)$$

with the constrained sums as before. From Eqs. (22) and (30) it can be seen that

$$\left. \frac{\partial^2 M}{\partial \theta_s \partial t} \right|_{\theta_s = \theta_i = \theta_r = 0} = \frac{d\langle S \rangle}{dt}, \quad (31)$$

with analogous expressions for $d\langle I \rangle / dt$ and $d\langle R \rangle / dt$.

Another way to obtain the above second derivative of M is by combining Eqs. (5), (11), and (14) to find $dP(s, i, r, t)/dt$. Then, multiplying dP/dt by $e^{\theta_s s + \theta_i i + \theta_r r}$, summing over s, r, i, and comparing with Eq. (31) yields

$$\frac{d\langle S \rangle}{dt} = -\alpha \langle SI \rangle,$$

$$\frac{d\langle I \rangle}{dt} = -\beta \langle I \rangle + \alpha \langle SI \rangle,$$

$$\frac{d\langle R \rangle}{dt} = \beta \langle I \rangle. \quad (32)$$

We note that the constraint $\langle S \rangle(t) + \langle I \rangle(t) + \langle R \rangle(t) = N$ at any t is compatible with Eq. (32). Further, in the large-N regime it is also interesting to observe that the mean-field-type approximation $\langle SI \rangle \approx \langle S \rangle \langle I \rangle$ turns [39] the system (32) identical to that of the conventional SIR model [1], consistent with the above discussion.

By solving Eq. (32) with the initial condition $\langle S \rangle(0) = N - 1$, $\langle I \rangle(0) = 1$, $\langle R \rangle(0) = 0$ in mean-field approximation through symbolic computation, we find the average subpopulations sizes and the basic reproduction number, respectively shown in circles in Figs. 6(a) and 6(b) for $N = 10^4$ (solid lines are to guide the eye in interpreting the overall behavior). Noticeably, these plots for large N show behaviors qualitatively similar to those for much smaller N, Fig. 2 and Fig. 5.

On the other hand, as discussed, the effect of fluctuations is central in smaller populations, implying significant differences between the results of the stochastic and deterministic SIR models in this regime. Clearly, however, this effect becomes gradually less important as N grows. For instance, in Appendix B we present the analytical expressions for R_0 with $N = 3, 4, 5$. If we write, e.g., the result for $N = 4$, Eq. (B4), as a function of the (not normalized) basic reproduction number of the *deterministic* SIR model [2], $R_0^{\text{det}} = N\alpha/\beta$, then the difference between R_0 and R_0^{det} can be readily inferred from the expression

$$R_0 = R_0^{\text{det}} \left[\frac{226}{5(3R_0^{\text{det}} + 8)} + \frac{2048}{7(3R_0^{\text{det}} + 16)} - \frac{2349}{70(R_0^{\text{det}} + 3)} \right.$$

$$\left. + \frac{18}{R_0^{\text{det}} + 4} - \frac{567}{5(R_0^{\text{det}} + 6)} + \frac{96}{5(R_0^{\text{det}} + 8)} \right]. \quad (33)$$

Moreover, if we consider R_0 for a larger $N = 20$ (whose expression is too cumbersome to be displayed), then it is also possible to determine the relative difference $(R_0^{\text{det}} - R_0)/R_0^{\text{det}}$, as presented in the heatmap plot of Fig. 7. Indeed, we note

FIG. 7. Heatmap plot displaying the relative difference $(R_0^{\text{det}} - R_0)/R_0^{\text{det}}$ between the basic reproduction number of the stochastic and deterministic SIR models for $N = 20$. Larger differences are noted in the important regime in which the infection rate surpasses the recovery rate, $\alpha \gtrsim \beta$.

FIG. 6. (a) Average numbers of susceptible ($\langle S \rangle$), infected ($\langle I \rangle$), and recovered ($\langle R \rangle$) individuals in a population of size $N = 10^4$ with $i_0 = 1$ initially infected case as a function of time t, for several infection rates α and fixed recovery rate $\beta = 1$. (b) Basic reproduction number $\mathcal{R}_0$ as a function of α, for fixed β values and $N = 10^4$. Despite the much larger population size, qualitative behaviors similar to those of Fig. 2 and Fig. 5 are noticed. In particular, when lower α values are considered the number of infectious cases decreases and reaches a maximum later in time (compare in (a) the solid and dashed red lines for $\alpha = 10$ and $\alpha = 2$, respectively). The solid lines in (b) are to guide the eye in interpreting the overall behavior.

that this difference becomes increasingly significant as the infection rate surpasses the recovery rate, $\alpha \gtrsim \beta$. Finally, by calculating R_0 via the Fock-space approach for progressively larger N, we find

$$R_0 = \left(\frac{N-1}{N} \right) f_N \left(\frac{R_0^{\text{det}}}{N} \right) R_0^{\text{det}}, \qquad (34)$$

where the function f_N is such that $f_N \to 1$ as $N \to \infty$, explicitly confirming the expected result that $R_0 \to R_0^{\text{det}}$ as $N \to \infty$.

V. CONCLUSIONS

The importance of improving the knowledge about the general dynamics of epidemic processes can be hardly overstate in current days. Over nearly a century SIR and related models

have been applied with this aim, since the seminal article by Kermack and McKendrick [1].

In this work we have addressed the stochastic SIR model in the Fock-space formalism, in which a master equation governs the transition probabilities among the system's states defined by SIR occupation numbers. This approach is particularly interesting for relatively small populations in which fluctuations not accounted for by conventional SIR models play a relevant role.

We have found for any population size N exact analytic expressions for the eigenvalues of the second-quantization Hamiltonian-like operator in Fock space that drives the epidemic infection and recovery processes. We have also presented small- and large-N results for the average subpopulations sizes and basic reproduction number as functions of the SIR model parameters α and β. For small N we have obtained the probability distributions of SIR states, epidemic sizes, and durations, which cannot be found from conventional SIR models based on ordinary differential equations for the populations sizes.

More generally, the Fock-space approach enables the derivation of higher moments and other measures of variation for stochastic systems, such as the individual-based SIR epidemiological model explored in this work, without the need for the stochastic simulation method or Monte Carlo approaches (for example, Refs. [7,49,50]). In particular, with such approaches the calculation of a likelihood, that is the probability of observed data conditioned on parameter values, requires extensive computation. In turn, this renders well-established techniques based on maximising the likelihood or Bayesian inference impractical, instead necessitating the extensive simulations required of likelihood free methods, such as approximate Bayesian computation [52]. In contrast, Fock-space methods render the calculation of the likelihood directly tractable, enabling immediate access to well-established and relatively efficient techniques for parameter inference, hypothesis testing and model selection, based on likelihood maximization or Bayesian inference. Thus the

current study in particular provides a framework for efficiently relating individual-based stochastic SIR models to observed data. Such features are of particular relevance to fitting, or selecting on, modeling predictions for variability arising from small number fluctuations, which cannot be captured by deterministic models, but are nonetheless particularly relevant in driving an initial epidemiological outbreak and the prospect of disease-variant emergence.

In addition, the derivation of analytical expressions such as the mean duration of an epidemic and the especially pertinent measure of whether disease will spread, the reproduction number, are presented for stochastic SIR systems for the first time to the best of our knowledge. Such expressions offer not only benchmarking for numerical algorithms of larger systems [40] but also fundamental insight into the how important features of disease dynamics may alter with stochasticity. Furthermore, a complete analytical characterisation of small stochastic epidemiological systems developed here offers the ability to efficiently incorporate large numbers of small stochastic subsystems, such as households and offices, within much larger city and national-scale epidemiological simulations.

We also remark that generalizations of the Fock-space approach to treat other epidemic models (e.g., SEIR, SITR, and SEQIJR) are readily feasible by changing the Hamiltonian and compartments in each case. Moreover, by including proper spatial constraints the study of the SIR stochastic lattice gas model could also be considered. Given the remarkable differences between the deterministic and stochastic SIR models in the small population regime, as well as the relevance of this regime at the beginning of an epidemic, these possibilities will be considered in forthcoming works.

We finally hope that the Fock-space approach discussed here can help to improve the understanding and characterization of the dynamics of epidemic processes, an issue that has become increasingly relevant in present days.

APPENDIX A: DERIVATION OF THE BASIC REPRODUCTION NUMBER $\mathcal{R}_0$ IN THE FOCK-SPACE APPROACH

Here we present a novel derivation of Eq. (29) for the basic reproduction number $\mathcal{R}_0$ in the context of the Fock-space approach to the stochastic SIR model.

We start by defining $z(t)$ as the average number of infected cases in time t due to a single infected individual that remained infectious since $t = 0$ (patient zero). Consider, also, the three following statistical events: Patient zero does not infect any susceptible individual in the time interval $[t, t + \Delta t]$ (E_0), patient zero infects only one individual in this interval (E_1), and more than one susceptible individual is infected by
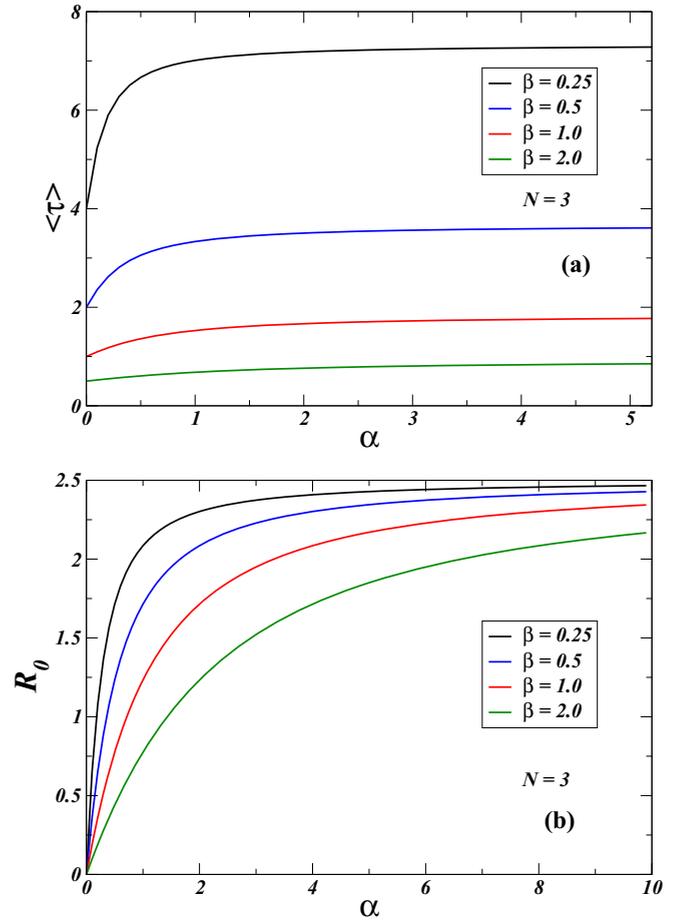
FIG. 8. (a) Mean epidemic duration $\langle \tau \rangle$ and (b) basic reproduction number $\mathcal{R}_0$ as a function of the infection rate α, for some fixed recovery rates β, and population size $N = 3$. Plots are taken from Eqs. (B2) and (B3), respectively. Behaviors similar to those of Fig. 4 and Fig. 5 are observed.

patient zero in this time interval ($E_{n>1}$). Here we take the infections of any two individuals as statistically independent events, so that the respective event probabilities must obey $P(E_0) + P(E_1) + P(E_{n>1}) = 1$ at any time t.

For sufficiently short time intervals (i.e., for $\Delta t \to dt$) the probability $P(E_{n>1})$ is negligible. So, from the definitions of the events above we write up to order $O(\Delta t)$,

$$z(t + \Delta t) = z(t)P(E_0) + [z(t) + 1]P(E_1). \quad \text{(A1)}$$

Recall from Eq. (8) that α and β are, respectively, the infection and recovery rates of the stochastic SIR model. So, in order to obtain $P(E_1)$ we have to multiply the average number $\langle S \rangle(t)$ of susceptible individuals in time t by the probability $\alpha \Delta t$ that one new infection occurs during this short time interval, and by the probability $P_{\text{nr}}(t)$ that patient zero has not been removed up to time t,

$$P(E_1) = \alpha P_{\text{nr}}(t) \langle S \rangle(t) \Delta t. \quad \text{(A2)}$$

We also note from Eq. (8) that $1 - \beta \Delta t$ gives the probability that an infected individual is not removed during this

sufficiently short time interval. We therefore write

$$P_{\text{nr}}(t + \Delta t) = P_{\text{nr}}(t)(1 - \beta \Delta t), \qquad (A3)$$

which in the limit $\Delta t \to 0$ becomes

$$\frac{dP_{\text{nr}}}{dt} = -\beta P_{\text{nr}}, \qquad (A4)$$

with solution for the initial condition $P_{\text{nr}}(0) = 1$ given by

$$P_{\text{nr}}(t) = e^{-\beta t}. \qquad (A5)$$

Now, by substituting Eq. (A5) into Eq. (A2) we find

$$P(E_1) = \alpha \langle S \rangle(t) \, \Delta t e^{-\beta t}. \qquad (A6)$$

Next, by using that $P(E_0) + P(E_1) = 1$ [up to order $O(\Delta t)$], we substitute $P(E_0)$ and $P(E_1)$ into Eq. (A1) and take the limit $\Delta t \to 0$, so that

$$\frac{dz}{dt} = \alpha e^{-\beta t} \langle S(t) \rangle. \qquad (A7)$$

Considering that patient zero is the only infected individual at $t = 0$ then $z(0) = 0$, and by integration

$$z(t) = \alpha \int_0^t e^{-\beta \tau} \langle S(\tau) \rangle \, d\tau. \qquad (A8)$$

The basic reproduction number then reads

$$\mathcal{R}_0 = \alpha \int_0^\infty e^{-\beta \tau} \langle S(\tau) \rangle \, d\tau, \qquad (A9)$$

which corresponds to Eq. (29) of the main text. We remark that this result has been previously derived [35] but in a distinct context using a different technique.

We can also use the eigenvectors expansion of the state vector, Eq. (21), along with Eq. (22) and $|s\,i\,r\rangle = \sum_\mu b_\mu |\lambda_\mu\rangle$, to obtain from Eq. (A9),

$$\mathcal{R}_0 = \alpha \sum_{s,i,r} s \sum_{\nu,\mu} a_\nu b_\mu \frac{\langle \lambda_\mu | \lambda_\nu \rangle}{\beta + \lambda_\nu}, \qquad (A10)$$

where $\langle \lambda_\mu |$ denotes the complex conjugate of the right eigenvector $|\lambda_\mu\rangle$ and the first sums are restricted to $s + i + r = N$, as usual.

APPENDIX B: CASE $N = 3$

We illustrate in this Appendix the calculation of explicit exact closed-form expressions for some important quantities of the stochastic SIR model in the Fock-space approach. Our aim is to show that this is actually feasible for small population sizes N. However, as larger N are considered the expressions become increasingly cumbersome.

In the case $N = 3$ the basis $\{|n\rangle = |s\,i\,r\rangle\}$ of the Fock space has dimension $N_h = (N + 1)(N + 2)/2 = 10$, comprising the following ordered set of pure Fock states:

$$\{|0, 0, 3\rangle, |0, 1, 2\rangle, |0, 2, 1\rangle, |0, 3, 0\rangle, |1, 0, 2\rangle,$$
$$|1, 1, 1\rangle, |1, 2, 0\rangle, |2, 0, 1\rangle, |2, 1, 0\rangle, |3, 0, 0\rangle\}.$$

From Eq. (15) we express the Hamiltonian-like matrix h on this basis as a function of the infection (α) and recovery (β) rates,

$$h = \begin{pmatrix}
0 & -\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \beta & -2\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2\beta & -3\beta & 0 & -\alpha & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 3\beta & 0 & 0 & -2\alpha & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\beta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \alpha+\beta & -2\beta & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 2(\alpha+\beta) & 0 & -2\alpha & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\beta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\alpha+\beta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}. \qquad (B1)$$

According to the discussion in Sec. III, the matrix h has $N + 1 = 4$ null eigenvalues $\lambda_\nu = 0$, associated with the eigenvectors $|\lambda_\nu\rangle = |N - \nu + 1, 0, \nu - 1\rangle = |4 - \nu, 0, \nu - 1\rangle$, for $\nu = 1, \ldots, 4$. The complete ordered set of exact eigenvalues is promptly obtained from Eqs. (18) and (19),

$$\{\lambda_\nu\} = \{0, 0, 0, 0, \beta, 2\beta, 3\beta, \alpha + \beta, 2(\alpha + \beta), 2\alpha + \beta\},$$

for $\nu = 1, \ldots, 10$, with corresponding right eigenvectors,

$$|\lambda_1\rangle = |3, 0, 0\rangle, \quad |\lambda_2\rangle = |2, 0, 1\rangle, \quad |\lambda_3\rangle = |1, 0, 2\rangle,$$
$$|\lambda_4\rangle = |0, 0, 3\rangle, \quad |\lambda_5\rangle = -|0, 0, 3\rangle + |0, 1, 2\rangle,$$
$$|\lambda_6\rangle = |0, 0, 3\rangle - 2|0, 1, 2\rangle + |0, 2, 1\rangle,$$
$$|\lambda_7\rangle = -|0, 0, 3\rangle + 3|0, 1, 2\rangle - 3|0, 2, 1\rangle + |0, 3, 0\rangle,$$

$$|\lambda_8\rangle = \frac{2\beta^2}{\beta^2 - \alpha^2}|0, 0, 3\rangle + \frac{2\beta}{\alpha - \beta}|0, 1, 2\rangle$$
$$+ \frac{\alpha}{\beta - \alpha}|0, 2, 1\rangle - \frac{\beta}{\alpha + \beta}|1, 0, 2\rangle,$$

$$|\lambda_9\rangle = \frac{\beta^3(5\alpha + 2\beta)}{(\alpha + \beta)^2(4\alpha^2 - \beta^2)}|0, 0, 3\rangle$$
$$+ \left[2\alpha\left(-\frac{1}{\alpha + \beta} + \frac{1}{2\alpha + \beta} + \frac{3}{\beta - 2\alpha}\right) + 4\right]|0, 1, 2\rangle$$
$$+ \frac{\beta(5\alpha + 2\beta)}{(2\alpha - \beta)(\alpha + \beta)}|0, 2, 1\rangle + \frac{2\alpha}{\beta - 2\alpha}|0, 3, 0\rangle$$
$$+ \frac{\beta^2}{(\alpha + \beta)^2}|1, 0, 2\rangle - \frac{2\beta}{\alpha + \beta}|1, 1, 1\rangle + |1, 2, 0\rangle,$$

$$|\lambda_{10}\rangle = \frac{2(5\alpha - 2\beta)\beta^2}{4\alpha^3 - 4\beta\alpha^2 - \beta^2\alpha + \beta^3}\,|0,0,3\rangle$$

$$+\left[-\frac{4\alpha}{\beta - 2\alpha} + \frac{6\alpha}{\beta - \alpha} + 4\right]|0,1,2\rangle$$

$$+2\alpha\left(\frac{1}{\beta - 2\alpha} + \frac{3}{\alpha - \beta}\right)|0,2,1\rangle$$

$$-\frac{2\alpha^2}{(\alpha - \beta)\beta}\,|0,3,0\rangle + \frac{4\beta}{2\alpha + \beta}\,|1,0,2\rangle - 4\,|1,1,1\rangle$$

$$+\frac{2\alpha}{\beta}\,|1,2,0\rangle + |2,0,1\rangle - \frac{\beta}{2\alpha + \beta}\,|2,1,0\rangle.$$

From the above sets of eigenvalues and eigenvectors we find, e.g., the mean duration of the epidemic and the basic reproduction number, respectively given in the Fock-space approach by Eqs. (27) and (A10),

$$\langle\tau\rangle = \frac{\alpha}{(\alpha + \beta)^2} + \frac{5}{6}\left(\frac{1}{2\alpha + \beta} - \frac{2}{\alpha + \beta}\right) + \frac{11}{6\beta}, \quad \text{(B2)}$$

$$R_0 = \alpha\left(\frac{27}{2\alpha + 3\beta} - \frac{8}{\alpha + 2\beta} - \frac{3}{\alpha + \beta}\right). \quad \text{(B3)}$$

Figures 8(a) and 8(b) display $\langle\tau\rangle$ and $\mathcal{R}_0$ calculated from Eqs. (B2) and (B3), respectively. We notice that these plots are qualitatively similar to Fig. 4 and Fig. 5, respectively, so that the associated discussion in Sec. IV.A also holds in the present case.

The above procedure can be readily applied to calculate significant quantities of the stochastic SIR model for smaller population sizes N. For example, by defining the ratio $\rho \equiv \alpha/\beta$, we list below the results for the basic reproduction number with $N = 4$ and $N = 5$, respectively,

$$R_0 = \frac{2\rho}{35}\left(\frac{315}{\rho + 1} + \frac{336}{\rho + 2} - \frac{3969}{2\rho + 3} + \frac{791}{3\rho + 2}\right.$$

$$\left. + \frac{5120}{3\rho + 4} - \frac{2349}{4\rho + 3}\right), \quad N = 4, \quad \text{(B4)}$$

$$R_0 = \rho\left[\frac{(129886\rho + 59213)}{1800(2\rho + 1)^2} - \frac{15}{\rho + 1} - \frac{80}{9(\rho + 2)}\right.$$

$$+ \frac{2349}{7(3 + 4\rho)} - \frac{9228}{35(2 + 3\rho)} + \frac{11583}{40(3 + 2\rho)}$$

$$\left. + \frac{78125}{84(4\rho + 5)} - \frac{192512}{175(4 + 3\rho)}\right], \quad N = 5. \quad \text{(B5)}$$

We mention that both expressions were obtained from *Mathematica* in about one minute of running time in a notebook with Intel Core i7 processor.

[1] W. O. Kermack and A. G. McKendrick, Proc. R. Soc. Lond. A **115**, 700 (1927).

[2] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications* (Griffin, London, 1975).

[3] D. S. Jones and B. D. Sleeman, *Differential Equations and Mathematical Biology* (Allen & Unwin, London, 1983).

[4] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1992).

[5] V. Capasso, *Mathematical Structure of Epidemic Systems* (Springer, Berlin, 1993).

[6] C. Nunn and S. Altizer, *Infectious Diseases in Primates: Behavior, Ecology and Evolution* (Oxford University Press, Oxford, 2006).

[7] M. Keeling, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ, 2007).

[8] E. Vynnycky and R. White, *An Introduction to Infectious Disease Modelling* (Oxford University Press, Oxford, 2010).

[9] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology* (Chapman & Hall, Boca Raton, FL, 2011).

[10] M. Martcheva, *An Introduction to Mathematical Epidemiology* (Springer, Berlin, 2015).

[11] M. S. Barlett, J. R. Statist. Soc. B **11**, 211 (1949).

[12] N. T. J. Bailey, Biometrika **37**, 193 (1950).

[13] D. G. Kendall, Proc. 3rd. Berkeley Symp. Math. Statist. Prob. **4**, 149 (1956).

[14] R. M. Anderson and R. M. May, Nature (Lond.) **280**, 361 (1979).

[15] R. M. Anderson, Bull. Math. Biol. **53**, 3 (1991).

[16] L. J. S. Allen, M. A. Jones, and C. F. Martin, Math. Biosci. **105**, 111 (1991).

[17] H. M. Hethcote, Math. Biosci. **145**, 89 (1997).

[18] M. J. Keeling, M. E. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell, Science **294**, 813 (2001).

[19] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[20] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).

[21] M. Boguná, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).

[22] M. J. Keeling, M. E. J. Woolhouse, R. M. May, G. Davies, and B. T. Grenfell, Nature (Lond.) **421**, 136 (2003).

[23] K. J. Sharkey, J. Math. Biol. **57**, 311 (2008).

[24] T. Britton, Math. Biosci. **225**, 24 (2010).

[25] D. Brockmann and D. Helbing, Science **342**, 1337 (2013).

[26] T. House, J. V. Ross, and D. Sirl, Proc. R. Soc. Lond. A **469**, 20120436 (2013).

[27] Y. Lin, D. Jiang, and P. Xia, Appl. Math. Comput. **236**, 1 (2014).

[28] A. J. Black and J. V. Ross, J. Theor. Biol. **367**, 159 (2015).

[29] C.-R. Cai, Z.-X. Wu, M. Z. Q. Chen, P. Holme, and J.-Y. Guan, Phys. Rev. Lett. **116**, 258301 (2016).

[30] I. Cooper, A. Mondal, and C. G. Antonopoulos, Chaos, Solitons Fractals **139**, 110057 (2020).

[31] F. Wong and J. J. Collins, Proc. Natl. Acad. Sci. U.S.A. **117**, 29416 (2020).

[32] T. Carletti, D. Fanelli, and F. Piazza, Chaos Solitons Fract. X **5**, 100034 (2020).

[33] S. Moein, N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, and Y. Gheisari, Sci. Rep. **11**, 1 (2021).

[34] B. F. Nielsen, L. Simonsen, and K. Sneppen, Phys. Rev. Lett. **126**, 118301 (2021).

[35] A. Ríos-Gutiérrez, S. Torres, and V. Arunachalam, Adv. Diff. Equ. **2021**, 288 (2021).

[36] W. Merbis and I. Lodato, Phys. Rev. E **105**, 044303 (2022).

[37] L. Mondaini, Biomed. Sci. Today **2**, e8 (2015).

[38] G. Jenkinson and J. Goutsias, PLoS One **7**, e36160 (2012).

[39] T. Tomé and M. J. de Oliveira, Braz. J. Phys. **50**, 832 (2020).

[40] M. Kröger and R. Schlickeiser, J. Phys. A: Math. Theor. **53**, 505601 (2020).

[41] R. Schlickeiser and M. Kröger, J. Phys. A: Math. Theor. **54**, 175601 (2021).

[42] C. Nguyen and J. M. Carlson, PLoS One **11**, e0152950 (2016).

[43] A. J. Black, T. House, M. J. Keeling, and J. V. Ross, J. R. Soc. Interface **10**, 20121019 (2013).

[44] E. C. Yuan, D. L. Alderson, S. Stromberg, and J. M. Carlson, PLoS One **10**, e0115826 (2015).

[45] C. S. Scherer, K. Adam, and R. T. N. Cardoso, Trends Comput. Appl. Math. **22**, 179 (2021).

[46] G. M. Schütz, M. Brandau, and S. Trimper, Phys. Rev. E **78**, 061132 (2008).

[47] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 2013).

[48] A. Amado, J. V. Santana-Filho, P. R. A. Campos, and E. P. Raposo, Eur. Phys. J. Plus **134**, 151 (2019).

[49] R. Erban, S. J. Chapman, and P. K. Maini, A practical guide to stochastic simulations of reaction-diffusion processes, arXiv:0704.1908.

[50] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).

[51] http://cain.sourceforge.net.

[52] F. Hartig, J. M. Calabrese, B. Reineking, T. Wiegand, and A. Huth, Ecol. Lett. **14**, 816 (2011).

[53] M. Doi, J. Phys. A: Math. Gen. **9**, 1465 (1976).

[54] M. Doi, J. Phys. A: Math. Gen. **9**, 1479 (1976).

[55] M. Schönberg, Il Nuovo Cimento **9**, 1139 (1952).

[56] M. Schönberg, Il Nuovo Cimento **10**, 419 (1953).

[57] P. T. Muzy, S. R. Salinas, A. E. Santana, and T. Tomé, Rev. Bras. Ensino Fis. **27**, 447 (2005).

[58] M. Sasai and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **100**, 2374 (2003).

[59] R. Dickman and I. Jensen, Phys. Rev. Lett. **67**, 2391 (1991).

[60] S. A. Isaacson, J. Phys. A: Math. Theor. **41**, 065003 (2008).

[61] F. Alcaraz, M. Droz, M. Henkel, and V. Rittenberg, Ann. Phys. **230**, 250 (1994).

[62] F. A. N. Santos, H. Gadêlha, and E. A. Gaffney, Phys. Rev. E **92**, 062714 (2015).

[63] G. C. Duarte-Filho, F. A. N. Santos, and E. A. Gaffney, Phys. Rev. E **102**, 052101 (2020).

[64] H. A. Araújo, M. O. Lukin, M. G. E. da Luz, G. M. Viswanathan, F. A. N. Santos, and E. P. Raposo, J. Stat. Mech. (2020) 083202.

[65] N. S. Nicolau, H. A. Araújo, G. M. Viswanathan, M. G. E. da Luz, and E. P. Raposo, J. Phys. A: Math. Theor. **54**, 325006 (2021).

[66] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, Cambridge, UK, 1999), chap. 6.

[67] T. Tomé and M. J. de Oliveira, *Stochastic Dynamics and Irreversibility* (Springer, Berlin, 2015), Chap. 13.

[68] R. Dickman, J. Stat. Phys. **55**, 997 (1989).

[69] I. Jensen and R. Dickman, J. Stat. Phys. **71**, 89 (1993).

[70] R. Dickman and R. Vidigal, Braz. J. Phys. **33**, 73 (2003).