

## Generalized Euler-Lotka equation for correlated cell divisions

Simone Pigolotti\*

*Biological Complexity Unit, Okinawa Institute of Science and Technology and Graduate University, Onna, Okinawa 904-0495, Japan*

(Received 11 January 2021; revised 1 April 2021; accepted 27 April 2021; published 30 June 2021)

Cell division times in microbial populations display significant fluctuations that impact the population growth rate in a nontrivial way. If fluctuations are uncorrelated among different cells, the population growth rate is predicted by the Euler-Lotka equation, which is a classic result in mathematical biology. However, cell division times can be significantly correlated, due to physical properties of cells that are passed through generations. In this Letter, we derive an equation remarkably similar to the Euler-Lotka equation which is valid in the presence of correlations. Our exact result is based on large deviation theory and does not require particularly strong assumptions on the underlying dynamics. We apply our theory to a phenomenological model of bacterial cell division in *E. coli* and to experimental data. We find that the discrepancy between the growth rate predicted by the Euler-Lotka equation and our generalized version is relatively small, but large enough to be measurable by our approach.

DOI: [10.1103/PhysRevE.103.L060402](https://doi.org/10.1103/PhysRevE.103.L060402)

Microbial populations in steady, nutrient-rich conditions tend to grow exponentially. Their exponential growth rate  $\Lambda$  can be taken as a proxy for the population fitness and is therefore a biologically important quantity. In a population of cells dividing at regular times  $\tau$ , the population size at a time  $T$  multiple of  $\tau$  is  $N(T) = N(0)2^{T/\tau}$ , so that  $\Lambda = \ln 2/\tau$ . In practice, cell division times of microbial populations significantly fluctuate, so that the division time  $\tau_i$  of a given individual  $i$  must be considered as a random quantity. As a consequence, the growth of  $N(T)$  is stochastic. In these situations, we can still define an exponential growth rate by

$$\Lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \ln N(T). \quad (1)$$

For independent, identically distributed cell division times  $\tau_i$ , the exponential growth rate converges to a deterministic value and can be computed as a solution of the celebrated Euler-Lotka equation

$$2 \langle e^{-\Lambda \tau} \rangle_\tau = 1, \quad (2)$$

where we denote the average over the distribution  $p(\tau)$  of the division times by  $\langle f(\tau) \rangle_\tau = \int d\tau p(\tau) f(\tau)$ . We use this notation also for discrete variables, with the integral appropriately replaced by a sum. Equation (2) is a classic result in mathematical biology. A recent experimental study has tested its prediction by tracking individual cell divisions in a microfluidic device [1]. Besides microbial populations, the Euler-Lotka equation finds important applications in epidemiology, where the factor 2 is replaced by the reproductive number  $R_0$  [2].

Experimental studies have revealed that fluctuations in microbial cell features are correlated among generations [3–5].

These correlations are caused by properties of cells that are passed from mother to daughter. These properties can be physical such as cell mass, or biological such as gene expression. Their fluctuations are often controlled to preserve homeostasis, i.e., a stable state of cells across generations. For example, experimental and theoretical studies provided evidences for an “adder” mechanism, in which cells attempt at growing their mass by a constant amount before dividing [4,6,7].

Regardless of the underlying mechanism, generalizing Eq. (2) to correlated cell divisions has proven to be a hard problem. One relatively simple case is the “Markovian” scenario where a cell division time conditionally depends only on that of her mother. Expressions for the growth rate in these cases have been derived in classic works by Powell [8] and Lebowitz and Rubinow [9] (see also [10,11]). Alternative approaches estimate the growth rate by comparing the outcome of sampling the population forward in time with retrospective sampling, in which individuals in the final population are traced back to their ancestors [12–15]. A recent study links the exponential growth rate  $\Lambda$  to the asymptotic distribution of the number of cell divisions  $\Delta$  among lineages [16] using large deviation theory. This approach has the advantage of neither requiring the Markovian assumption, nor retrospective sampling.

In this Letter, we introduce a generalized Euler-Lotka equation [Eq. (2)] which is valid for correlated cell division times:

$$\lim_{\Delta \rightarrow \infty} \frac{1}{\Delta} \ln \left\langle e^{-\Lambda \sum_{i=1}^{\Delta} \tau_i} \right\rangle_{\{\tau_i\}} = -\ln 2, \quad (3)$$

where  $\langle \dots \rangle_{\{\tau_i\}}$  denotes an average over sequences  $\{\tau_i\} = \tau_1, \tau_2, \dots, \tau_\Delta$  of cell division times in independent lineages. If the  $\tau_i$ s were uncorrelated, then the left-hand side of Eq. (3) reduces to the cumulant generating function  $\ln \langle e^{q\tau} \rangle_\tau$ , and therefore Eq. (3) becomes equivalent to the traditional Euler-Lotka equation [Eq. (2)]. Equation (3) only requires as

\*simone.pigolotti@oist.jp

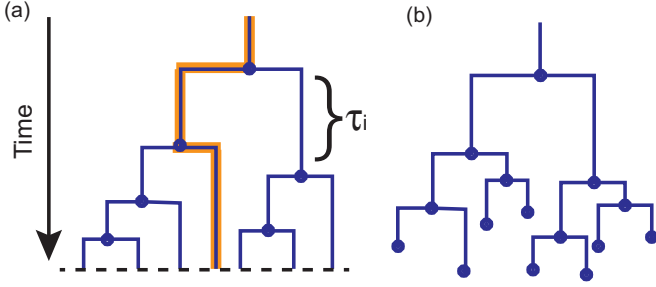


FIG. 1. Population dynamics represented as a lineage tree. (a) A microbial population grows in time from a single cell. Nodes (circles) denote cell division events. Lengths of branches denote the cell division times  $\tau_i$ s. One lineage is represented with a thick orange line. In this case, the population is let to evolve until a fixed time  $T$ . (b) Lineage tree in an alternative ensemble, in which lineages are let to evolve until they have accumulated exactly  $\Delta = 4$  cell divisions.

hypotheses that population dynamics is at steady state and the sum of the  $\tau_i$ s across a lineage satisfies a large deviation principle with a convex rate function, which in practice are rather mild assumptions (see, e.g., Secs. 3.5.6 and 4.4 of Ref. [17]). Equation (3) can therefore be used to compute the population growth rate from individual lineages in rather general settings.

We consider a microbial population initially constituted of a single individual. The population grows in time by a sequence of cell divisions. We represent the genealogy of the population by a tree, whose nodes are cell division events and branches are times between consecutive cell divisions [see Fig. 1(a)].

We now introduce the concept of a lineage. A lineage is identified by an individual in the population at time  $T$  complemented by its past history, i.e., the number  $\Delta$  of cell divisions separating it from the individual at time zero and the sequence of cell division times  $\{\tau_i\}$  of all its ancestors [see Fig. 1(a)]. Following Refs. [16,18], we now imagine to randomly select a lineage by starting from the initial individual and picking at each division one of the two newborns with equal probability. With this procedure, a lineage that includes  $\Delta$  cell divisions is chosen with probability  $2^{-\Delta}$ . We approximate the probability that this randomly selected lineage includes  $\Delta$  cell division events by the empirical frequency  $p(\Delta; T) \approx 2^{-\Delta} N(\Delta; T)$ , where  $N(\Delta; T)$  is the number of lineages with  $\Delta$  cell divisions at time  $T$ . Since  $N(T) = \sum_{\Delta} N(\Delta; T)$ , we obtain

$$N(T) \approx \langle 2^{\Delta} \rangle_{\Delta}. \quad (4)$$

Substituting this expression into the definition of the exponential growth rate, Eq. (1), we find

$$\Lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \langle 2^{\Delta} \rangle_{\Delta}. \quad (5)$$

At variance with Eq. (4), Eq. (5) is an exact equality, as the empirical frequencies  $2^{-\Delta} N(\Delta; T)$  converge to  $p(\Delta; T)$  in the limit  $T \rightarrow \infty$  [16,19].

To make further progress, we introduce some ideas from large deviation theory [17]. Large deviation theory describes the leading behavior of probability distributions when a parameter (like the time  $T$  in our case) becomes large. In large deviation theory, variables such as  $\Delta$ , whose average is pro-

portional to  $T$ , are called extensive. We associate with  $\Delta$  the intensive variable  $\delta = \Delta/T$ , whose average tends to a constant for large  $T$ . The large deviation principle for  $\delta$  is expressed by

$$p(\delta) \asymp e^{-T I^{(\delta)}(\delta)}. \quad (6)$$

The function  $I^{(\delta)}(\delta)$  is called the rate function. We use the notation  $I^{(\delta)}$  to stress that  $I$  is the rate function associated with the distribution of the variable  $\delta$ . The symbol “ $\asymp$ ” denotes the leading exponential behavior; it can be seen as a shorthand for  $I(\delta) = -\lim_{T \rightarrow \infty} [\ln p(\delta)]/T$ .

An alternative way of studying asymptotic fluctuations of intensive random variables is via the scaled cumulant generating function, defined by

$$\psi^{(\delta)}(q) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \langle e^{qT\delta} \rangle_{\delta}. \quad (7)$$

The Gartner-Ellis theorem states that, if the rate function is convex, it is related with the scaled cumulant generating by a Legendre-Fenchel transform

$$I^{(\delta)}(\delta) = \sup_q [q\delta - \psi^{(\delta)}(q)]. \quad (8)$$

Since the Legendre-Fenchel transform is an involution, it also holds that  $\psi^{(\delta)}(q) = \sup_{\delta} [q\delta - I^{(\delta)}(\delta)]$ .

We now return to Eq. (4) and briefly summarize the main result of Ref. [16]. Assuming that  $\delta$  satisfies a large deviation principle, we obtain

$$\Lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \int d\delta e^{T[\delta \ln 2 - I^{(\delta)}(\delta)]}. \quad (9)$$

In the limit  $T \rightarrow \infty$ , the integral can be evaluated with the method of steepest descent, obtaining

$$\Lambda = \sup_{\delta} [\delta \ln 2 - I^{(\delta)}(\delta)]. \quad (10)$$

Equation (10) is the central result of Ref. [16]. An alternative way to obtain it is to directly identify the expression of the scaled cumulant generating function in Eq. (5):

$$\Lambda = \psi^{(\delta)}(\ln 2). \quad (11)$$

Equation (10) then follows by expressing the scaled cumulant generating function in terms of the rate function by means of the Gartner-Ellis theorem.

Application of this theory requires knowledge of the asymptotic distribution of  $\Delta$ , or its intensive counterpart  $\delta$ . However, in analogy with the Euler-Lotka equation [Eq. (2)], it would be desirable to express  $\Lambda$  in terms of the distribution of division times and its correlations. To this aim, we consider a case in which, rather than letting the population grow until a given time  $T$ , each lineage is let to grow until it has accumulated exactly  $\Delta$  cell divisions [see Fig. 1(b)]. In this alternative *ensemble*,  $\Delta$  is fixed whereas  $T$  fluctuates among lineages. In this case, we consider  $T$  as an extensive random variable, since its average grows linearly with the fixed large parameter  $\Delta$ . We similarly associate with  $T$  the intensive variable  $t = T/\Delta$ . We expect  $t$  to satisfy as well a large deviation principle:

$$p(t) \asymp e^{-\Delta I^{(t)}(t)}. \quad (12)$$

In the language of probability theory and in particular of queuing theory,  $\Delta(T)$  is called a counting process and  $T(\Delta)$  its inverse. A useful result [20] states that the rate functions of their associated intensive variables,  $\delta$  and  $t$ , respectively, are related by

$$I^{(\delta)}(x) = xI^{(t)}(1/x). \quad (13)$$

Note that this is the result that one would obtain by simply applying the rule for a change of variable in the large deviation form of the probability distribution.

We now substitute this result into Eq. (10), obtaining

$$\Lambda = \sup_{\delta} \left\{ \delta \left[ \ln 2 - I^{(t)}\left(\frac{1}{\delta}\right) \right] \right\}, \quad (14)$$

and, by applying the Gartner-Ellis theorem,

$$\begin{aligned} \Lambda &= \sup_{\delta} \left\{ \delta \left[ \ln 2 - \sup_q \left( \frac{q}{\delta} - \psi^{(t)}(q) \right) \right] \right\} \\ &= \sup_{\delta} \inf_q [\delta \ln 2 - q + \delta \psi^{(t)}(q)]. \end{aligned} \quad (15)$$

We assume that the function in square brackets smoothly depends on  $\delta$  and  $q$  and therefore compute the supremum and infimum by simply taking derivatives. The extremality condition with respect to  $\delta$  is expressed by

$$\psi^{(t)}(q^{\text{inf}}) = -\ln 2. \quad (16)$$

Substituting this condition back into Eq. (15) yields  $\Lambda = -q^{\text{inf}}$ , so that we rewrite Eq. (16) as

$$\psi^{(t)}(-\Lambda) = -\ln 2. \quad (17)$$

Upon substituting the definition of the scaled cumulant generating function, Eq. (7), into Eq. (17), we obtain the generalized Euler-Lotka equation [Eq. (3)], as anticipated.

Taking the derivative in Eq. (15) with respect to  $q$  results in

$$\delta^{\max} \frac{d}{dq} \psi^{(t)}(q) \Big|_{q=-\Lambda} = 1. \quad (18)$$

This equation relates the dominant value of  $\delta$  with the statistics of the division times and provides another facet to the generalized Euler-Lotka theory. Equation (18) is best interpreted in the simple case of uncorrelated cell divisions, where it reduces to  $\delta^{\max} = \langle \tau e^{-\Lambda \tau} \rangle^{-1}$ . If  $\Lambda \ll 1$ , the dominant value of  $\gamma$  is simply its average value, i.e., the inverse of the average division time. However, for quickly growing population, the dominant value of  $\delta$  becomes significantly larger than this value, as cells that reproduce faster contribute more to population growth.

To illustrate our result, we consider a phenomenological model of bacterial growth inspired by the adder principle [4]. In the model, each bacterial cell grows in length at a rate  $\alpha$ . The rate  $\alpha$  is generated by the formula

$$\alpha = \alpha_0 + c(\alpha_M - \alpha_0) + \sigma_{\alpha}(1 - c^2)\xi, \quad (19)$$

where  $\alpha_M$  is the value of  $\alpha$  of the mother of the considered cell. The parameters  $\alpha_0$  and  $\sigma_{\alpha}$  are the average and variance of the distribution of  $\alpha$ , respectively. The variable  $\xi$  is a Gaussian random variable with zero average and unit variance. The parameter  $c$  controls the degree of correlations between the growth rate of mothers and daughters.

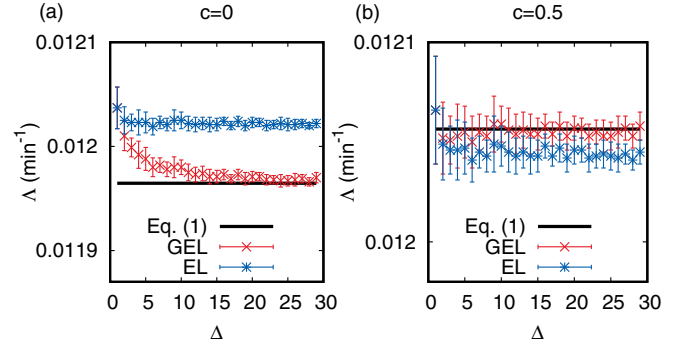


FIG. 2. Estimates of the growth rate  $\Lambda$  in the adder model of Ref. [4] obtained by the direct definition given in Eq. (1), by the generalized Euler-Lotka equation (3) (GEL), and by the conventional Euler-Lotka equation (2) (EL). In both panels, parameters of the growth rate distribution are  $\alpha_0 = 0.0255 \text{ min}^{-1}$  and  $\sigma_{\alpha} = 0.0027 \text{ min}^{-1}$ . The inherited length fraction  $f$  is distributed according to a Gaussian with mean  $f_0 = 0.5$  and standard deviation  $\sigma_f = 0.03$ . The added length  $l$  follows a lognormal distribution with mean  $l_0 = 3.21 \mu\text{m}$  and standard deviation  $\sigma_l = 0.54 \mu\text{m}$ . (a) Growth rates of mothers and daughters are uncorrelated ( $c = 0$ ). (b) Growth rates of mothers and daughters are positively correlated ( $c = 0.5$ ). In both panels, we average over  $n_{\text{lin}} = 5000$  lineages and plot the results as a function of the number of cell divisions  $\Delta$  in each lineage. In all panels, simulations are repeated 20 times; error bars denote standard deviations computed from these realizations. Details on the direct numerical estimate of  $\Lambda$  are presented in the Supplemental Material [21].

Each cell is characterized by a length  $s_b$  at birth and  $s_d$  at death. The adder model postulates that the added length  $l = s_d - s_b$  is roughly constant among cells. After division, a daughter inherits a fraction  $f$  of the mother's length. We allow for some variability by taking both  $f$  and  $l$  as random variables, whose distributions are estimated from experimental data [4]. The time between cell divisions is expressed by

$$\tau = \frac{1}{\alpha \ln(s_d/s_b)} = \frac{1}{\alpha \ln(1 + l/s_b)}. \quad (20)$$

To understand the dynamics of this model, we first consider the simple case in which the growth rate  $\alpha$  is uncorrelated across generations,  $c = 0$ . In this case, the adder dynamics causes *negative* correlations in cell division times. An intuitive explanation is that cells that grow to a large size tend to have a long generation time and to give birth to larger cells. These larger cells, in turn, tend to have a short generation time because of the dependence of  $\tau$  on  $s_b$  in Eq. (20). Since the traditional Euler-Lotka equation neglects correlations, we expect it to overestimate  $\Lambda$  in this case. This idea is confirmed in Fig. 2(a), which also shows how the generalized Euler-Lotka equation correctly estimates the directly measured value of  $\Lambda$ .

We now move to a case in which growth rates are correlated across generations,  $c = 0.5$ . This value is compatible with observation from *E. coli* experiments [4]. In this case, the positive correlations induced by the growth rate tend to counterbalance those caused by the adder model. As a result, both the traditional and the generalized Euler-Lotka equations

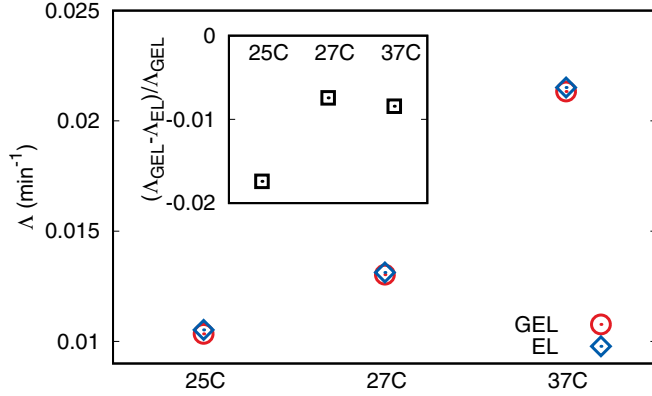


FIG. 3. Growth rate  $\Lambda$  estimated from Eq. (2) and Eq. (17) from the experiments in [22,23]. The experimental data consist of a number of lineages equal to 65, 54, and 160 for three different temperatures 25°C, 27°C, and 37°C, respectively. Each of these lineages includes 70 division events. For the EL estimate, we consider all division events at each temperature. For the GEL estimate, we use all sublineages of length  $\Delta = 20$  from all lineages at each temperature. The inset shows the relative difference between the two estimates.

predict a growth rate that is close to the correct one, although the traditional Euler-Lotka equation slightly underestimates it [see Fig. 2(b)].

We now test the generalized Euler-Lotka equation on experimental data from [22,23]. These experiments tracked a large number of *E. coli* lineages at different temperatures. Applying our approach, we find that the generalized Euler-Lotka equation predicts a lower growth rate than the traditional Euler-Lotka equation (see Fig. 3). The differences are on the order of 1% and suggest that the compensatory mechanisms observed in Ref. [4] are less effective in the experiments by [22,23]. A direct inspection of the correlation functions of the cell division times supports this idea (see the Supplemental Material [21]).

An advantage of our approach is that it allows us to use the arsenal of techniques from large deviation theory [17] to compute the scaled cumulant generating function, and thereby the growth rate via Eq. (3). For example, we apply our theory to the Markovian case in which the division time conditionally depends on the maternal division time only, i.e.,

$$p(\tau_i | \tau_{i-1}, \tau_{i-2}, \tau_{i-3}, \dots) = p(\tau_i | \tau_{i-1}). \quad (21)$$

In this case, one has

$$\langle e^{q \sum_{i=1}^{\Delta} \tau_i} \rangle = \int d\tau_1 \cdots d\tau_{\Delta} p_{\tau_1} e^{q\tau_1} \prod_{i=2}^{\Delta} e^{q\tau_{i-1}} p(\tau_i | \tau_{i-1}). \quad (22)$$

Equation (22) expresses the generating function as an iterated convolution. This iterated convolution is characterized by the integral kernel  $K(\tau_i, \tau_{i-1}) = e^{q\tau_{i-1}} p(\tau_i | \tau_{i-1})$ . For large  $\Delta$ , the behavior of the left-hand side of Eq. (22) is dominated by the leading eigenvalue  $\lambda(q)$  of its integral kernel  $K$ . In particular, the definition of the scaled cumulant generating function, Eq. (7), implies that

$$\psi^{(\tau)}(q) = \ln \lambda(q). \quad (23)$$

Combining Eq. (3) and Eq. (23) we obtain  $2\lambda(-\Lambda) = 1$ . Calling  $\phi(\tau)$  the eigenvector associated with this eigenvalue, we express this result as

$$\phi(\tau) = 2 \int d\tau' e^{-\Lambda\tau'} p(\tau | \tau') \phi(\tau'). \quad (24)$$

Equation (24) is a classic result for the Markovian case [8,9,24]. The eigenvector  $\phi(\tau)$  can be interpreted as the distribution of division times that one would measure over the whole tree, including cells that have not divided yet at time  $T$  [11].

In conclusion, in this paper we derived the generalized Euler-Lotka equation [Eq. (3)]. This equation describes the growth rate of populations where cell divisions occur in a correlated way. We obtained this result by means of a result in queuing theory [20] that was recently applied in stochastic thermodynamics [25] and to study enzymes replicating information [26]. We have demonstrated that our result can be easily applied to lineage data. A comparison with the prediction of the traditional Euler-Lotka equation permits to quantitatively assess the impact of correlations on the population growth rate. Due to these properties, we expect the generalized Euler-Lotka equation to become a useful tool to analyze lineages in experimental population dynamics.

I thank Deepak Bhat and Anzhelika Koldaeva for discussions. I am grateful to Arthur Genthon, David Lacoste, Luca Peliti, and two anonymous referees for feedback on the manuscript.

- 
- [1] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, *Proc. Natl. Acad. Sci. USA* **113**, 3251 (2016).
  - [2] N. C. Grassly and C. Fraser, *Nat. Rev. Microbiol.* **6**, 477 (2008).
  - [3] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, *Curr. Biol.* **20**, 1099 (2010).
  - [4] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, *Cell* **159**, 1433 (2014).
  - [5] M. Wallden, D. Fange, E. G. Lundius, Ö. Baltekin, and J. Elf, *Cell* **166**, 729 (2016).
  - [6] A. Amir, *Phys. Rev. Lett.* **112**, 208102 (2014).
  - [7] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, *Curr. Biol.* **25**, 385 (2015).
  - [8] E. Powell, *Microbiology* **15**, 492 (1956).
  - [9] J. L. Lebowitz and S. Rubinow, *J. Math. Biol.* **1**, 17 (1974).
  - [10] J. Lin and A. Amir, *Cell Syst.* **5**, 358 (2017).
  - [11] E. Levien, J. Kondev, and A. Amir, *J. R. Soc. Interface* **17**, 20190827 (2020).
  - [12] T. J. Kobayashi and Y. Sughiyama, *Phys. Rev. Lett.* **115**, 238102 (2015).
  - [13] R. García-García, A. Genthon, and D. Lacoste, *Phys. Rev. E* **99**, 042413 (2019).

- [14] A. Genthon and D. Lacoste, *Sci. Rep.* **10**, 11889 (2020).
- [15] A. Genthon and D. Lacoste, [arXiv:2012.02734](https://arxiv.org/abs/2012.02734).
- [16] E. Levien, T. GrandPre, and A. Amir, *Phys. Rev. Lett.* **125**, 048102 (2020).
- [17] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [18] T. Nozoe, E. Kussell, and Y. Wakamoto, *PLoS Genet.* **13**, e1006653 (2017).
- [19] F. Jafarpour, C. S. Wright, H. Gudjonson, J. Riebling, E. Dawson, K. Lo, A. Fiebig, S. Crosson, A. R. Dinner, and S. Iyer-Biswas, *Phys. Rev. X* **8**, 021007 (2018).
- [20] P. W. Glynn and W. Whitt, *Queueing Syst.* **17**, 107 (1994).
- [21] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.103.L060402> for details on the direct numerical estimate of the exponential growth rate. The Supplemental Material also presents the correlation functions for the model of Fig. 2 and for the experimental data of Fig. 3.
- [22] Y. Tanouchi, A. Pai, H. Park, S. Huang, R. Stamatov, N. E. Buchler, and L. You, *Nature (London)* **523**, 357 (2015).
- [23] Y. Tanouchi, A. Pai, H. Park, S. Huang, N. E. Buchler, and L. You, *Sci. Data* **4**, 170036 (2017).
- [24] J. Lin and A. Amir, *Phys. Rev. E* **101**, 012401 (2020).
- [25] T. R. Gingrich and J. M. Horowitz, *Phys. Rev. Lett.* **119**, 170601 (2017).
- [26] D. Chiuchiu, Y. Tu, and S. Pigolotti, *Phys. Rev. Lett.* **123**, 038101 (2019).