Universality in Stochastic Exponential Growth

Srividya Iyer-Biswas,1 Gavin E. Crooks,2 Norbert F. Scherer,1,* and Aaron R. Dinner1,†
1James Franck Institute and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, USA
2Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
(Received 17 December 2013; published 7 July 2014)

Recent imaging data for single bacterial cells reveal that their mean sizes grow exponentially in time and that their size distributions collapse to a single curve when rescaled by their means. An analogous result holds for the division-time distributions. A model is needed to delineate the minimal requirements for these scaling behaviors. We formulate a microscopic theory of stochastic exponential growth as a Master Equation that accounts for these observations, in contrast to existing quantitative models of stochastic exponential growth (e.g., the Black-Scholes equation or geometric Brownian motion). Our model, the stochastic Hinshelwood cycle (SHC), is an autocatalytic reaction cycle in which each molecular species catalyzes the production of the next. By finding exact analytical solutions to the SHC and the corresponding first passage time problem, we uncover universal signatures of fluctuations in exponential growth and division. The model makes minimal assumptions, and we describe how more complex reaction networks can reduce to such a cycle. We thus expect similar scalings to be discovered in stochastic processes resulting in exponential growth that appear in diverse contexts such as cosmology, finance, technology, and population growth.

DOI: 10.1103/PhysRevLett.113.028101 PACS numbers: 87.17.Ee, 05.40.-a, 87.18.Tt

Discovering unifying physical principles that transcend the complexity of specific biological systems is a fundamental goal of the field of biological physics [1,2]. Quantitative analyses of gene regulatory networks have revealed general connections between network motifs, fluctuations in the dynamics of participating molecules, and biological functions at the molecular scale [3–6]. Analogous quantitative relationships governing behaviors at the organismal scale are just beginning to emerge [1,7,8]. In particular, in a recent experiment, we found that scaling laws governed the stochastic growth of individual Caulobacter crescentus cells [9]. In the same study, the sizes of the cells were shown to increase exponentially between divisions, consistent with observations for other microorganisms [10–13].

Exponential growth is ubiquitous and has been studied in diverse contexts [11,14]. It describes inflation of the universe, geometric multiplication of an entity of interest (e.g., nuclear or cellular fission), and phenomenological dynamics (e.g., the Black-Scholes equation for options prices; Moore’s law for computer processor power). Many such processes are inherently stochastic, with the times between contributing events drawn from waiting-time distributions [15]. Surprisingly, given its prevalence, there is no microscopic model for stochastic exponential growth. While various other physical aspects of cell growth have been examined previously [16–20], a theory relating the statistics of the stochastic exponential growth to essential features of the biochemical networks underlying growth is needed.

A phenomenological model of stochastic exponential growth, a Langevin equation with linear drift and linear multiplicative noise, was famously applied by Black and Scholes to explain financial data on stock options prices; it forms the basis of modern quantitative derivative trading [21]. This model, also known as geometric Brownian motion (GBM), has since been used extensively in various cellular contexts, and when applied to cell growth, it predicts a log-normal cell size distribution [22]. However, in this model, the standard deviation grows faster than the exponentially growing mean such that the ratio, i.e., the coefficient of variation (COV), increases as the square root of time. This prediction is in disagreement with observations in [9], wherein the COV of cell sizes was found to be time invariant.

Here, we provide a microscopic theory of stochastic exponential growth that yields the universality of fluctuations during the growth of single bacterial cells, observed in [9]; it also agrees with the aforementioned observed constancy of the COV with time. This microscopic theory is built on the assumption that growth is governed by an autocatalytic cycle of reactions. We argue a posteriori that this is the minimal model consistent with the observations in [9]. Furthermore, we provide a theoretical framework for examining stochastic cell division and show how scale invariance of division time distributions arises. We also discuss why the essential features of the model are retained even when more complex network topologies govern cell growth.

Stochastic Hinshelwood cycle.—Our stochastic theory builds on a simple deterministic (kinetic) model introduced in 1952 by Hinshelwood [23]. In this model components of the cell that govern cell growth are connected through an
autocatalytic cycle of reactions in which each species catalyzes the production of the next [Fig. 1(a)]. The mass (or equivalently, the size) of a cell is assumed to be proportional to a linear combination of the copy numbers of the species in the cycle. We construct a stochastic generalization of this cycle, by assuming that the waiting times for the individual reactions are exponentially distributed, i.e., that the reactions are elementary. We refer to this model as the stochastic Hinshelwood cycle (SHC). In general, the SHC contains \( N \) species, \( \{X_1, X_2, \ldots, X_N\} \). The scaling laws that we derive do not depend on their identities or \( N \). The mean rate of production of \( X_i \) is \( k_i x_{i-1} \), where \( x_{i-1} \) is the copy number of \( X_{i-1} \) (Fig. 1). For use below, we write this rate as a matrix multiplication: 

\[
k_{x_{i-1}} = \sum_{j=1}^{N} k_{ij} x_j,
\]

where \( k \) is the rate constant matrix with elements \( k_{ij} = k_i \delta_{i-1,j} \), and \( \delta \) is the Kronecker delta. In this notation, the reaction scheme is

\[
X_{i-1} \xrightarrow{k_{i-x_{i-1}}} X_i + X_{i-1},
\]

for \( 1 < i, j < N \); the index 0 is equivalent to \( N \), closing the cycle.

We denote the state of the general \( N \)-step SHC model by the vector \( x \equiv (x_1, x_2, \ldots, x_N) \), where \( x_i \) is the copy number of \( X_i \) present at a given time. The corresponding Chemical Master Equation [15,24] for the time evolution of the probability distribution, \( P(x,t) \), is

\[
\frac{\partial P}{\partial t} = \sum_{i,j=1}^{N} k_{ij} x_j [P(x_i-1, \ldots, x_j, \ldots, x_N) - P(x_i, \ldots, x_j, \ldots, x_N)].
\]

(2)

From (2), we derive the time evolution equations for the moments of \( x \) from the eigenvalues and eigenvectors for \( k \). Since \( k \) is a cyclic matrix of period \( N \), \( k^N = k \). The eigenvalues of the rate constant matrix are the \( N \) complex roots of unity times \( \kappa \), the geometric mean of all the rates. The \( m \)th eigenvalue is \( \kappa_m = \kappa \exp(i 2 \pi m/N) \), and the \( q \)th component of the corresponding eigenvector is \( \xi^{(q)}_m = (\prod_{p=1}^N k_p)/\kappa^q_m \). In the asymptotic limit (i.e., when \( t \gg 1/\kappa \)),

\[
\mu_q(t) \sim \sum_{i=1}^{N} \text{U}_{q} \text{U}_{N}^{\dagger} \mu_i(0) e^{\kappa t}.
\]

(3)

Thus, the mean copy numbers of all reactants evolve asymptotically as \( e^{\kappa t} \). Moreover, the dependence on initial conditions for \( \mu_q(t) \) is independent of \( q \). It follows that the ratio of any two mean copy numbers \( \mu_q(t)/\mu_r(t) \) is equal to \( \text{U}_{q_N}/\text{U}_{r_N} \), which is independent of initial conditions and depends only on the \( q \)th and \( r \)th components of the \( N \)th eigenvector, \( \xi_N \).

**Time evolution of growth fluctuations.**—To examine the time evolution of growth fluctuations, we determine the equation of motion of the covariance matrix \( C_{ij} \equiv \langle (x_i x_j) - \langle x_i \rangle \langle x_j \rangle \rangle \) [15]. In matrix form,

\[
\frac{d}{dt} C(t) = [k] C(t) + \text{C}(t) [k]^{\dagger} + \frac{d}{dt} \Xi(t),
\]

(4)

where \( [k] \) denotes the transpose and \( \Xi(t) \) is an \( N \times N \) diagonal matrix with entries \( \xi_i(t) = \delta_{ij} \mu_j(t) \). We have computed the exact analytical solution for the time evolution of the covariance matrix [25]. In the asymptotic limit,

\[
C_{ij}(t) \sim \sum_{p=1}^{N} b_p \mu_p(0) e^{2 \kappa t},
\]

(5)

where \( b_p \) is a coefficient that depends only on the \( p \)th and not the initial conditions [25]. Thus, \( C_{ij} \) scales as \( e^{2 \kappa t} \) for all \( i \) and \( j \). Moreover, the time-independent prefactor of element \( C_{ij} \) of the covariance matrix is proportional to \( \text{U}_{iN} \text{U}_{jN}^{\dagger} \). Combining (5) with (3) gives
FIG. 2 (color online). Copy number fluctuations are perfectly correlated in the asymptotic limit. (a) Ratios of the copy numbers of the components of the $N = 3$ SHC from the trajectories shown in Fig. 1(b). As predicted by (6), for $t \gg 1/\kappa$, the ratios of the different $x_i$ tend to constant values in each ensemble member, a signature of the perfect correlations between component copy numbers in the asymptotic state. (b) Emergence of a single composite variable. The variable $s_i$ is the projection of the state vector $x$ onto the $i$th eigenvector $(\xi_i)$ of the rate constant matrix, $K$. We see that $s_3$ here (or more generally, $s_n$) tends to a constant nonzero level, while the remainder of the projections vanish.

\[
\text{Cov}[x_i(t)/\mu_i(t), x_j(t)/\mu_j(t)]/\sigma_i \sigma_j \sim 1, \tag{6}
\]

where $\sigma_i$ is the standard deviation of the rescaled variable $x_i/\mu_i(t)$.

An important consequence of (6) is that asymptotically all $x_i$ are proportional to each other, since two stochastic variables can be perfectly correlated only when they are linearly related [15]. Thus, the ratio of any two of them must asymptote to a time-independent constant value in each ensemble member [i.e., each cell; see Fig. 2(a)], but this value itself has a distribution across different members. We note that $x_i(t)$ and $x_j(t)$ themselves continue to fluctuate in each stochastic realization even as their ratio tends to a constant value.

**Scalings of the size distribution.**—Two different scaling laws are encapsulated in (6). First, every rescaled variable $x_i/\mu_i(t)$ has the same distribution in the asymptotic limit. Second, since $e^{\kappa t}$ is a scaling variable, the distribution shape for each $x_i$ is invariant with time, even as its mean increases exponentially. Therefore, the $n$th moment of $x_i$ goes as $e^{n\kappa t}$.

For clarity, we explicitly compute the size distribution for the case when all rate constants in the model are equal, with value $\kappa$. In this case, $K$ becomes a circulant matrix, and the projection of the state vector $x$ onto the asymptotically dominant eigenvector $\xi_N$ reduces to a simple sum of the constituent copy numbers, $s \equiv \sum_{i=1}^{N} x_i$. This variable $s$ itself undergoes dynamics governed by a $N = 1$ SHC. Then, $P(s, t)$ for the initial condition $P(s, t = 0) = \delta_{s, s_0}$, is the negative binomial distribution,

\[
P(s, t|s_0, 0) = \binom{s \kappa t}{s - s_0} e^{(s_0 - 1) \kappa t} (1 - e^{-\kappa t})^{s - s_0}. \tag{7}
\]

This result can be verified by direct substitution into (2). In the continuum limit for $s$, (7) tends to a gamma distribution, since the negative binomial distribution can be written as a Poisson mixture of gamma distributions [26]. Asymptotically,

\[
P(s, t \to \infty|s_0, 0) = \frac{s_0^{s_0 - 1} e^{-(s_0 \kappa t)}}{s_0^s \Gamma(s_0)} . \tag{8}
\]

For the general case with unequal rates, the analog of $s$ is the linear combination of $\{x_i\}$ that is defined by the projection of the state vector along the eigenvector corresponding to the largest eigenvalue, $K$: $s_N \equiv \sum_{i=1}^{N} U_{iN} x_i$. As shown in Fig. 2(b), all $s_i \equiv \sum_{j=1}^{N} U_{ij} x_j$ for $j \neq N$ vanish in the long-time limit, and the only contributions to fluctuations in each $x_i$ come from $s_N$. As a result, all $x_i/\mu_i$ are distributed in the same fashion as $s$ in (8) (Fig. 3), with $s_0 = s_N(0)$. In other words, the mean-rescaled distribution of cell sizes must fit the same gamma distribution at all times.

**Division as the first passage time to a size threshold.**—We assume that cell division occurs when the cell size $s$ reaches a threshold [9,13]. In general, this threshold can be absolute ($s$ itself attains a critical value), relative ($s$ increases by a critical multiple), or differential ($s$ increases by a critical amount). In the absence of additional feedback mechanisms, the scaling derived above implies that the different components of the SHC maintain their predivision.
provided that \( \lambda \) are the eigenvalues of the characteristic polynomial that determines the (complex) order of the polynomial is \( N \).

Substituting (8) into (9), we find that the first passage time distribution from a given initial size \( s_0 \) to an absolute threshold \( \theta \) is a beta-exponential distribution [27],

\[
P(\tau) = \frac{\kappa e^{-s_0 \pi \tau} (1 - e^{-\pi \tau})^{\theta - s_0}}{B(s_0, 1 + \Theta - s_0)},
\]

where \( B \) is the beta function. The first passage time distributions for differential or relative size thresholds can be found using this expression.

**Scalings of division times.**—Since \( \tau \) always occurs as \( \kappa \tau \) in (10), we can rescale time by \( \kappa^{-1} \), or, equivalently, by \( \pi(\tau) \), to obtain a universal scale-invariant shape of the division time distribution. A complementary translational collapse of \( P(\tau) \) is obtained when \( \tau \) is shifted to \( \tau - \log(\theta/\kappa) \), provided that \( \theta \gg s_0 \). The scale invariance of first passage time distributions is more universal than (10); a similar scaling collapse of the division time distribution will be obtained whenever a single time scale dominates the dynamics, regardless of the thresholding scheme (i.e., absolute, relative, or differential). Operationally, this implies that if \( \kappa \) is varied by changing an external parameter (e.g., nutrient quality, oxygen concentration, osmotic pressure, or temperature), the mean-rescaled division time distributions for different values of \( \kappa \) should collapse to the same curve, as observed in [9].

**Extensions of the SHC model.**—More complex autocatalytic network topologies can be specified by augmenting \( \kappa \) in (1) by additional nonzero entries. In this case, the characteristic polynomial that determines the (complex) eigenvalues \( \lambda \) of the augmented reaction matrix is [28]

\[
\sum_{\text{cycles}} \left( \frac{\kappa_{\text{cycle}}}{\lambda} \right)^{N_{\text{cycle}}} = 1.
\]

In other words, a complex autocatalytic network can be factorized into irreducible cycles, each with \( N_{\text{cycle}} \) members [29]. \( \kappa_{\text{cycle}} \) is the geometric mean of rates of a given cycle. Since there is always one cycle with all \( N \) members, the order of the polynomial is \( N \). The largest \( \kappa_{\text{cycle}} \) determines which cycle dominates the asymptotic dynamics; the linked members of that cycle specify an effective SHC, and remaining species entrain to its stochastic exponential growth dynamics. Thus, all the SHC scaling predictions continue to hold for more complex topologies [28].

**Discussion.**—In this Letter, we have introduced the stochastic Hinshelwood cycle (Fig. 1), a model of stochastic exponential growth. Its dynamics naturally lead to the emergence of a single composite growth variable with a single time scale, thus yielding scaling collapses for size and division time distributions (Figs. 2 and 3), as observed in [9]. Moreover, this model explains the observed Arrhenius scaling of the exponential growth rate [9,30–32]: since the effective exponential growth rate is the geometric mean of the individual rates, the effective activation energy barrier is the arithmetic mean of the individual ones, and, thus, of the order of a single enzyme reaction’s, i.e., \( \approx 13 \) kcal/mol [9].

Unlike GBM, the SHC model predicts that the ratio of the standard deviation to the mean (COV) of cell sizes is asymptotically a constant, in agreement with observations in [9]. This can be directly seen from (6), or by writing down the phenomenological Langevin description corresponding to the SHC: \( ds/dt = \kappa s(t) + \sqrt{s(t)} \eta(t) \) (\( \eta \) is standard delta-correlated Gaussian white noise), whose solution is the gamma distribution in (8) [33,34]. In contrast, GBM has a noise term \( s(t) \eta(t) \) in the variables above and results in a log-normal size distribution [22] [35] with an asymptotically diverging COV \(( \propto \sqrt{t})\).

The differences in the predictions of the two models (SHC vs GBM) have important implications. In favorable chemostatic conditions, bacterial cells grow exponentially at a constant rate. The single-cell analog of this “balanced growth condition” is that the mean-rescaled cell-size distributions remain invariant, even as the cells grow and divide. This has been observed in [9], and is obtained from the SHC but not GBM. In the SHC the scaling collapse of cell-size distributions reflects the statistical self-similarity of the underlying stochastic process, which ensures constancy of COV.

The success of the SHC raises the question of its molecular origin. As discussed, a complex autocatalytic network governing cell growth can be systematically reduced to an effective SHC, with an exponential growth rate determined by a subset of connections. Moreover, the mechanics of cell wall synthesis must be coupled to cell growth via the regulation of number density of active growth sites by a component of the SHC [37]. Previous studies have found indirect evidence that there are two key steps governing bacterial growth: the global production of proteins at a rate proportional to the numbers of ribosomal RNA and vice versa [31,32]—in essence, an \( N = 2 \) stochastic Hinshelwood cycle [11]. It would be interesting to test these ideas directly by designing perturbations that give rise to \( N \)-dependent transients.
We thank Tom Witten and Leo Kadanoff for several insightful discussions, and Herman Gudjonson for a careful reading of the manuscript. A. R. D., N. F. S., and S. I. B. thank the W. M. Keck Foundation and the National Science Foundation (NSF Grant No. PHY-1305542) for financial support. G. E. C. was supported by the Office of Basic Energy Sciences of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

\[ \text{nfschere@uchicago.edu} \\
\text{dinner@uchicago.edu} \]