Gibbs free energy change of a discrete chemical reaction event (EP) FREE

Carlos Floyd (iD) ; Garegin A. Papoian ✉ (iD) ; Christopher Jarzynski ✉ (iD)

Check for updates

 CHORUS

CrossMark

View Online

Export Citation

Articles You May Be Interested In

Gibbs entropy and dynamics

Chaos (May 2008)

Gibbs vs Boltzmann Entropies

American Journal of Physics (May 1965)

Gibbs states for AF algebras

J. Math. Phys. (December 1998)

AIP Publishing

Gibbs free energy change of a discrete chemical reaction event ᴱᴾ

View Online   Export Citation   CrossMark

Carlos Floyd,[1] 🆔 Garegin A. Papoian,[1,2,3,a] 🆔 and Christopher Jarzynski[1,2,3,4,a] 🆔

AFFILIATIONS

[1] Biophysics Program, University of Maryland, College Park, Maryland 20742, USA

[2] Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, USA

[3] Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA

[4] Department of Physics, University of Maryland, College Park, Maryland 20742, USA

[a] Authors to whom correspondence should be addressed: gpapoian@umd.edu and cjarzyns@umd.edu

ABSTRACT

In modeling the interior of cells by simulating a reaction–diffusion master equation over a grid of compartments, one employs the assumption that the copy numbers of various chemical species are small, discrete quantities. We show that, in this case, textbook expressions for the change in Gibbs free energy accompanying a chemical reaction or diffusion between adjacent compartments are inaccurate. We derive exact expressions for these free energy changes for the case of discrete copy numbers and show how these expressions reduce to traditional expressions under a series of successive approximations leveraging the relative sizes of the stoichiometric coefficients and the copy numbers of the solutes and solvent. Numerical results are presented to corroborate the claim that if the copy numbers are treated as discrete quantities, then only these more accurate expressions lead to correct behavior. Thus, the newly derived expressions are critical for correctly computing entropy production in mesoscopic simulations based on the reaction–diffusion master equation formalism.

I. INTRODUCTION

In recent years, the coarse-grained computational modeling of intracellular environments has enjoyed significant advances. An important paradigm shared by many such models is to treat the evolution of reacting chemical species' copy numbers and spatial distributions by simulating a reaction–diffusion master equation (RDME).[1] In this approach, the system volume is divided into compartments, each with local values of the copy numbers and chemical potentials of the different chemical species (Fig. 1). The RDME is a differential equation describing the evolution of the probability $P(\mathbf{N}, t)$ of observing the vector of copy numbers $\mathbf{N} = \{N_{i,A}\}_{i \in L, A \in \Omega}$ of chemical species i in compartment A at time t, where L is the set of solute species and Ω is the set of compartments in the system. The RDME can be written schematically as

$$\frac{dP(\mathbf{N}, t)}{dt} = (\hat{M} + \hat{D})P(\mathbf{N}, t), \tag{1}$$

where $\hat{M}$ and $\hat{D}$ represent operators describing chemical reactions and inter-compartment diffusion, respectively.[2] An in-depth description of the RDME approach can be found in Refs. 3 and 4, where the forms of the operators are discussed. Rather than directly solving Eq. (1), one often simulates trajectories of the vector $\mathbf{N}$ obeying the stochastic dynamics encoded in the RDME using a variant of the Gillespie algorithm.[5] The sizes of the compartments are commonly determined by the Kuramoto lengths, the mean free diffusional path for a species before it participates in a chemical reaction. Within each compartment, the spatial distributions of the reacting species are assumed to be homogeneous, allowing the use of mass-action kinetics with the compartment's local values of the species' concentrations to describe the stochastic chemical reaction propensities. Molecules can additionally jump between adjacent compartments in "diffusion events" (whose propensities also depend on the compartments' local concentrations of species) to give rise to concentration gradients on the scale of the compartment
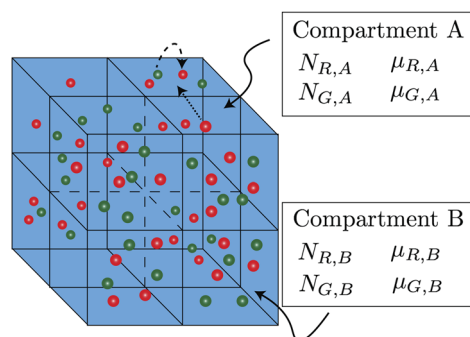
FIG. 1. An example of a cubic compartment grid used in the simulation of a RDME. Each compartment, labeled with letters A, B, . . ., has local values of the quantities N_R, N_G, referring to the copy numbers of the red and green molecules in the compartment, and of μ_R, μ_G, referring to the chemical potentials of those molecules. Molecules can react with each other within compartments (long dashed arrow), as well as hop between adjacent compartments, representing diffusion (short dashed arrow).

length. This modeling approach is appropriate when the Kuramoto length is small compared to the system size (i.e., the assumption of homogeneity over the system volume fails), yet large compared to the intermolecular distance scale.[1] Examples of simulation platforms based on such an approach include Virtual Cell,[6,7] lattice microbes,[8,9] MesoRD,[10] MEDYAN (Mechanochemical Dynamics of Active Networks),[11,12] and others.[13–17]

One important aspect of simulating non-equilibrium biological systems is the computation of thermodynamic forces that drive the observed flux on the network of chemical reactions.[18–20] Determining these forces can allow for the quantification of entropy production in chemically reactive systems.[21,22] In several research groups, measuring entropy production in biological active matter has been a recent goal.[23,24] For instance, in recent work, we quantified the entropy production rates of self-organizing non-equilibrium actomyosin networks in MEDYAN using the expressions derived here as a first step.[12] The ability to measure dissipation in active matter systems will allow us to test the applicability of different physical organizing principles relating the production of entropy to the likelihood of observing certain trajectories.[25,26] For isothermal, isobaric, chemically reactive solutions, which include most mesoscopic biological systems, measuring the total entropy production amounts to determining the change in Gibbs free energy that accompanies chemical reactions and diffusion down concentration gradients.[27–29] A ubiquitous textbook expression for the change in Gibbs free energy G accompanying a chemical reaction is

$$\Delta G = k_B T \log K_{eq} + k_B T \log Q, \tag{2}$$

where K_{eq} is the equilibrium constant, Q is the reaction quotient (we give definitions of these quantities below), k_B is Boltzmann's constant, and T is the temperature.[30–35] At equilibrium, $Q = K_{eq}^{-1}$, and as a result, $\Delta G = 0$. However, in this paper, we argue that Eq. (2) is a biased approximation to the exact value of ΔG accompanying a chemical reaction that holds when the copy numbers of the reacting molecules are large, such as on the order of Avogadro's number. When the system is small, such as when copy numbers are on the

order of 100 as is often the case in RDME simulations of intracellular environments, thermodynamic expressions such as Eq. (2) require corrections.[36,37]

As a simple motivating example, consider a mixture of an even total number of two chemical species, red and green, which interconvert at equal rates. At equilibrium, the copy numbers of these molecules will be equal, and $Q = K_{eq}^{-1} = 1$. Now, if a reaction were to occur at equilibrium to produce one additional red molecule and to eliminate one green molecule. Then we would expect that the Gibbs free energy of the system had increased, since we have left equilibrium where the free energy attains its minimum. However, the prediction of Eq. (2) is that $\Delta G = 0$ for this reaction. The assumption whose violation leads to Eq. (2) being incorrect is that the copy number of chemical species is a continuous quantity. When these variables are considered as discrete, a different expression for ΔG must be used to give correct behavior.

Similarly, for diffusion between adjacent compartments, a common expression for the change in Gibbs free energy accompanying the jump of a molecule i from compartment A with copy number $N_{i,A}$ to compartment B, where its copy number is $N_{i,B}$, is

$$\Delta G = k_B T \log \frac{N_{i,B}}{N_{i,A}}. \tag{3}$$

Imagine, however, we have a situation where $N_{i,A} = N_{i,B}$ and a molecule jumps from compartment A to B. The Gibbs free energy should have increased since we have departed from the highest entropy distribution of the molecules over the two compartments; however, Eq. (3) will predict that $\Delta G = 0$.

In this paper, we derive exact expressions for the change in Gibbs free energy accompanying chemical reactions within compartments and diffusion events between compartments, and we further show how these expressions relate to the familiar textbook formulas [Eqs. (2) and (3)] through a series of approximations. We also discuss the assumptions involved in defining the Gibbs free energy of a grid of homogeneously mixed compartments that can exchange energy and particles, such as that used in a simulation of a RDME. Finally, we present numerical simulations using MEDYAN to demonstrate the need to use these more exact expressions for ΔG in order to obtain sensible results when copy numbers are treated as discrete variables. Only these more exact expressions will give correct, unbiased behavior when measuring entropy production in mesoscopic *in silico* studies of biological non-equilibrium systems that rely on the RDME formalism.

II. METHODS

A. ΔG of reactions

Here, we make successive approximations to the formula for ΔG accompanying a chemical reaction, and our notation reflects the level of approximation in which certain quantities are being used: when appropriate, we subscript quantities with a parenthesized number, i.e., $\Delta G_{(0)}$, where increasing numbers represent more approximate versions. The symbol $\sim$ will indicate that the quantities of chemical species are being represented by mole fractions χ_i, rather than concentrations C_i. In this section, we treat the case

that our system comprises a single closed compartment of a homogeneous dilute solution in which a chemical reaction has occurred, and we derive an expression for ΔG. In this system, the number of solvent molecules is fixed and the solute molecules participate in chemical reactions, causing their copy numbers to change. In Secs. II B and II C, we consider a system with multiple weakly interacting compartments (which may, in general, comprise an irregular grid rather than the regular Cartesian grid visualized in Fig. 1), each of which contains a homogeneous solution with local copy numbers of solvent and solutes and between which both solvent and solutes can diffuse. The nearly exact result [Eq. (21)] obtained in this section will also apply to those systems, as we argue below.

Before restricting to the case of a single closed compartment, we establish notation for properties of the chemical species in a compartment grid. The chemical potential for species i in compartment A can be expressed as depending either upon the mole fraction, $\chi_{i,A}$, or upon the concentration, $C_{i,A}$, of that species in the compartment,

$$\mu_{i,A} = \widetilde{\mu_i^0}(T,p) + k_B T \log \chi_{i,A} = \mu_i^0(T,p) + k_B T \log C_{i,A}, \quad (4)$$

where k_B is Boltzmann's constant, $\widetilde{\mu_i^0}(T,p)$ is the standard state chemical potential at temperature T and pressure p when working with the dimensionless $\chi_{i,A}$, and $\mu_i^0(T,p)$ is the same when working with $C_{i,A}$. $C_{i,A}$ and $\chi_{i,A}$ both play the role of the "composition variable" leading to these different, yet equivalent expressions for the chemical potential.[38] We make the distinction between dependence upon copy number and dependence upon concentration in order to establish parameters that can be used in simulation, which commonly works with copy numbers, based on those given in the literature, which typically use units of concentration. Here, we make the assumption of an ideal-dilute solution and neglect the coefficient of activity of the solute species.[39] The mole fraction can be written as

$$\chi_{i,A} = \frac{N_{i,A}}{N_A} = \frac{N_{i,A}}{\sum_{j \in M} N_{j,A}}, \quad (5)$$

where N_A is the total copy number of molecules in compartment A and M is the set of all species including the solvent in the system. Similarly, the concentration can be written as

$$C_{i,A} = \frac{N_{i,A}}{\Theta_A}, \quad (6)$$

where

$$\Theta_A = N_{Av} V_A \quad (7)$$

is a conversion factor, N_{Av} is Avogadro's number, and V_A is the compartment volume (we assume constant pressure and that the fluctuations in volume are negligible for these liquid systems, allowing the use of Gibbs free energy). Using Eqs. (4)–(6), we can write[40]

$$\widetilde{\mu_i^0}(T,p) = \mu_i^0(T,p) + k_B T \log \frac{N_A}{\Theta_A}. \quad (8)$$

Using standard arguments concerning thermodynamic extensivity, it is possible to establish that the Gibbs free energy of the solution in compartment A can be written as a weighted sum over the chemical potentials of the species,

$$G_A(\mathbf{N}_A) = \sum_{i \in M} N_{i,A} \mu_{i,A}(\chi_{i,A}), \quad (9)$$

where $\mathbf{N}_A = \{N_{i,A}\}_{i \in M}$ is the vector of species copy numbers $N_{i,A}$, and where we have explicitly indicated the dependency of $\mu_{i,A}$ upon mole fraction $\chi_{i,A}$ via Eq. (5).[28,41] One may be concerned that Eq. (9) fails to apply when the copy numbers of solutes are small. As explained below, we make the assumption that boundary effects are still negligible, which allows us to treat G as a first order homogeneous function of the number of copies of the system.[37] This is the necessary property to establish Eq. (9), so the small copy numbers of solutes do not render this approach invalid. We rely on Eqs. (4) and (9) to derive changes in Gibbs free energy accompanying chemical reactions and inter-compartment diffusion.

Consider a reaction of the general form

$$v_1 X_1 + v_2 X_2 + \cdots \underset{k_-}{\overset{k_+}{\rightleftharpoons}} v_1 Y_1 + v_2 Y_2 + \cdots, \quad (10)$$
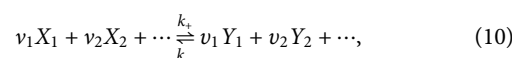
where X_i represent reactants, Y_j represent products, v_i and v_j are stoichiometric coefficients, and the rate of reaction is k_+ to the right and k_- to the left. We have dropped the subscript A indicating the compartment in which the reaction takes place and now restrict to the case that our system is a single compartment. When this reaction has occurred once to the right, the copy numbers of reactants have changed $N_i \rightarrow N_i - v_i$ and those of the products have changed $N_j \rightarrow N_j + v_j$. We calculate the change in Gibbs free energy accompanying this reaction by considering it as resulting from these finite, discrete changes in copy numbers,[37] not from infinitesimal changes. Using Eqs. (4), (5), and (9), the Gibbs free energy before the reaction has occurred can be written as

$$G^{\text{initial}} = \sum_{i \in R} N_i \left(\widetilde{\mu_i^0} + k_B T \log \frac{N_i}{N} \right) + \sum_{j \in P} N_j \left(\widetilde{\mu_j^0} + k_B T \log \frac{N_j}{N} \right) + N_s \left(\widetilde{\mu_s^{0*}} + k_B T \log \frac{N_s}{N} \right), \quad (11)$$

where R is the set of reactants, P is the set of products, the subscript s refers to the solvent, and where we have dropped the dependence of the standard state chemical potential on T and p. $\widetilde{\mu_i^0}$ and $\widetilde{\mu_j^0}$ describe the chemical potential at a reference concentration of the solute in the solvent (also referred to as the solute standard state), whereas $\widetilde{\mu_s^{0*}}$ describes the chemical potential at a reference state of pure solvent (also referred to as the solvent standard state).[39] We assume here, for simplicity and without loss of generality, that there are no solute species that have not participated in the reaction (i.e., spectator solute species). These species would also contribute terms to Eq. (11), but when we subtract the initial from the final Gibbs free energy, their inclusion would not yield a different result. The final Gibbs free energy is

$$G^{\text{final}} = \sum_{i \in R} (N_i - v_i) \left(\widetilde{\mu_i^0} + k_B T \log \frac{N_i - v_i}{N + \sigma} \right) + \sum_{j \in P} (N_j + v_j) \left(\widetilde{\mu_j^0} + k_B T \log \frac{N_j + v_j}{N + \sigma} \right) + N_s \left(\widetilde{\mu_s^{0*}} + k_B T \log \frac{N_s}{N + \sigma} \right), \quad (12)$$

where

$$\sigma = \sum_{j \in P} v_j - \sum_{i \in R} v_i, \tag{13}$$

which is the amount by which the total species copy number N has changed. As described in Refs. 38 and 42, when $\sigma \neq 0$, it is important to account for the solvent species in the calculation of free energy differences. This is because the free energy of the solvent, which is the most abundant species in the reaction volume, will not be the same after the reaction has taken place since its mole fraction will change as N changes to $N + \sigma$. Neglecting the solvent species when $\sigma \neq 0$ leads to expressions for ΔG that are off by an amount $\sigma k_B T$.[38] Whereas the authors of Refs. 38 and 42 describe the appearance of this erroneous term while formulating the Gibbs free energy as a function of a continuous degree of advancement of reaction $d\xi = -dN_i/v_i = dN_j/v_j$, here we treat the extent of reaction as a discrete quantity. In the limit that $v_i/N_i \to 0$, $v_j/N_j \to 0$ for each reactant and product, the discrete case passes into the continuum case; however, under the assumption of small copy numbers, we do not satisfy this limit. In Appendix A, we discuss further differences between the continuum treatment and the discrete treatment, as well as the relation to the Gibbs–Duhem equation.

After some algebra (using the fact that $\sum_{i \in R} N_i + \sum_{j \in P} N_j + N_s = N$), we can write the change in Gibbs free energy as

$$\Delta G_{(0)} = G^{\text{final}} - G^{\text{initial}} = \widetilde{\Delta G^0} + k_B T \log \prod_{i \in R} \frac{(N_i - v_i)^{N_i - v_i}}{N_i^{N_i}}$$
$$\times \prod_{j \in P} \frac{(N_j + v_j)^{N_j + v_j}}{N_j^{N_j}} \frac{N^N}{(N + \sigma)^{N+\sigma}}, \tag{14}$$

where

$$\widetilde{\Delta G^0} = \sum_{j \in P} v_j \widetilde{\mu_j^0} - \sum_{j \in R} v_i \widetilde{\mu_i^0}. \tag{15}$$

Equation (14) is exact, but we would like to avoid specifying N in simulation since the solvent is typically not modeled explicitly, as we elaborate on in Sec. II B. We would also like to determine $\widetilde{\Delta G^0}$ from literature values of ΔG^0. To these ends, we first rewrite Eq. (14) as

$$\Delta G_{(0)} = \widetilde{\Delta G^0} + k_B T \log \widetilde{Q}_{(0)} + k_B T \log \frac{N^N}{(N + \sigma)^{N+\sigma}}, \tag{16}$$

where

$$\widetilde{Q}_{(0)} = \prod_{i \in R} \frac{(N_i - v_i)^{N_i - v_i}}{N_i^{N_i}} \prod_{j \in P} \frac{(N_j + v_j)^{N_j + v_j}}{N_j^{N_j}}. \tag{17}$$

From Eqs. (8) and (15), we can write

$$\widetilde{\Delta G^0} = \Delta G^0 + \sigma k_B T \log \frac{N}{\Theta}, \tag{18}$$

where

$$\Delta G^0 = \sum_{j \in P} v_j \mu_j^0 - \sum_{i \in R} v_i \mu_i^0. \tag{19}$$

Inserting this into Eq. (16), we have

$$\Delta G_{(0)} = \Delta G^0 + k_B T \log \left(\frac{N}{\Theta}\right)^\sigma \frac{N^N}{(N + \sigma)^{N+\sigma}} + k_B T \log \widetilde{Q}_{(0)}$$
$$= \Delta G^0 - \sigma k_B T \log \Theta + k_B T \log \left(\frac{N}{N + \sigma}\right)^{N+\sigma} + k_B T \log \widetilde{Q}_{(0)}. \tag{20}$$

We now make the approximation that $N \gg \sigma$, which is certainly reasonable for most realistic parameterizations of the compartment grid (in the example of a 0.125 μm^3 compartment filled with water, $N \sim 10^9$ while $\sigma \sim 1$). With this, we can write the third term in Eq. (20) as $-\sigma k_B T$,[43] giving

$$\Delta G_{(1)} = \Delta G^0 - \sigma k_B T \log \Theta - \sigma k_B T + k_B T \log \widetilde{Q}_{(1)}, \tag{21}$$

where $\widetilde{Q}_{(1)} = \widetilde{Q}_{(0)}$ (we updated the subscript to indicate the use of this quantity in a more approximate version of the formula for ΔG). We recommend using Eq. (21) in simulation because it allows us to incorporate literature values for ΔG^0 and avoids specification of N. To understand the term $-\sigma k_B T$ in Eq. (21) and to understand the relationship between Eqs. (21) and (16) and the textbook expression for ΔG, we proceed by making further approximations leveraging the large sizes of the solute copy numbers compared to their stoichiometric coefficients. First, we rewrite $\widetilde{Q}_{(1)}$ as

$$\widetilde{Q}_{(1)} = \prod_{i \in R} \left(1 - \frac{v_i}{N_i}\right)^{N_i} (N_i - v_i)^{-v_i} \prod_{j \in P} \left(1 + \frac{v_j}{N_j}\right)^{N_j} (N_j + v_j)^{v_j}. \tag{22}$$

Assuming $N_i \gg v_i$ and $N_j \gg v_j$, and using the limit

$$\lim_{x \to \infty} \left(1 \pm \frac{y}{x}\right)^x = e^{\pm y}, \tag{23}$$

we obtain

$$k_B T \log \widetilde{Q}_{(1)} \approx \sigma k_B T + k_B T \log \prod_{i \in R} (N_i - v_i)^{-v_i} \prod_{j \in P} (N_j + v_j)^{v_j}. \tag{24}$$

Inserting this into Eq. (21), canceling the term $\sigma k_B T$, gives

$$\Delta G_{(2)} = \Delta G^0 - \sigma k_B T \log \Theta + k_B T \log \widetilde{Q}_{(2)}, \tag{25}$$

where

$$\widetilde{Q}_{(2)} = \prod_{i \in R} (N_i - v_i)^{-v_i} \prod_{j \in P} (N_j + v_j)^{v_j}. \tag{26}$$

Finally, introducing

$$\widetilde{Q}_{(3)} = \prod_{i \in R} N_i^{-v_i} \prod_{j \in P} N_j^{v_j} \tag{27}$$

and discarding terms in $\Delta G_{(2)}$ that scale like $\frac{v_i}{N_i}$ or $\frac{v_j}{N_j}$, we arrive at

$$\Delta G_{(3)} = \Delta G^0 - \sigma k_B T \log \Theta + k_B T \log \widetilde{Q}_{(3)} = \Delta G^0 + k_B T \log Q, \tag{28}$$

where

$$Q = \prod_{i \in R} C_i^{-v_i} \prod_{j \in P} C_j^{v_j}. \tag{29}$$

We see that Eq. (28) is obtained from Eq. (25) upon making the approximations $N_i - v_i \approx N_i$ and $N_j + v_j \approx N_j$. This final result in

Eq. (28) is a standard textbook expression for the change in Gibbs free energy.[30–35]

We, thus, have four expressions for ΔG of chemical reactions:

- Equation (14) for $\Delta G_{(0)}$ is exact; however, it requires specifying N.
- Equation (21) for $\Delta G_{(1)}$ uses the approximation $N \gg \sigma$. We recommend the use of this expression because it is the most exact expression for which we need not specify N, and since it is written in terms of ΔG^0, for which literature values can be found.
- Equation (25) for $\Delta G_{(2)}$ uses the approximations $N_i \gg v_i$ and $N_j \gg v_j$.
- Equation (28) for $\Delta G_{(3)}$ uses the approximations $N_i \gg v_i$ and $N_j \gg v_j$ again.

In Appendix B, we provide expressions for the accuracy of these approximations. Note that, without specifying the copy number of solvents N_s, we can only approximately compute changes in Gibbs free energy, not the instantaneous Gibbs free energy of the system. Typically, only the changes are of interest.

The definition of Gibbs free energy states that

$$\Delta G = \Delta H - T\Delta S. \qquad (30)$$

Comparing this expression to Eq. (28), we identify

$$\Delta G^0 = \Delta H - T\Delta S^0 \qquad (31)$$

and

$$-T\Delta S_{\text{mix}} = k_B T \log Q, \qquad (32)$$

where ΔH is the enthalpy of reaction, and ΔS^0 and ΔS_{mix} represent the changes in entropy due to the molecular conformations and the translational motions, respectively,

$$\Delta S = \Delta S^0 + \Delta S_{\text{mix}}. \qquad (33)$$

It is instructive to realize that the discrepancies between the various expressions for ΔG are due to how the ΔS_{mix} term in Eq. (33) is written.

When employing any of the above expressions that involve the logarithms of products of copy numbers that are raised to the power of other copy numbers, we recommend splitting the logarithm of products into a sum of logarithms as well as using $\log x^y = y \log x$ in order to prevent overflow resulting from computing very large numbers. The results of this section do not assume a system consisting of multiple, weakly interacting compartments, which we discuss next.

B. Thermodynamics of a reaction–diffusion compartment grid

Now that we have treated the scenario of a reaction event occurring within a single compartment, we want to generalize to the case of a grid of compartments. For this, we develop an argument based on timescales that will allow us to track the copy numbers of the reactive solutes while ignoring those of the inert solvent. This is a necessary modeling feature due to the computational infeasibility of tracking the solvent copy numbers in each compartment for long times. After describing the thermodynamic framework for a grid of compartments, in Sec. II C, we derive the change in G resulting from diffusion of solutes between adjacent compartments.

In simulating a RDME, one commonly treats diffusion between adjacent compartments and chemical reactions within compartments using an augmented set of all species and reactions in the system that treats species as distinct if they belong to separate compartments. Thus, if there are $|L|$ reacting species and $|\Omega|$ compartments, where L and Ω are the sets of solute species and compartments, respectively, then in the augmented set, there are $|L\|\Omega|$ species tracked. The number of reactions in the augmented system, including r chemical reactions per compartment and roughly $z|L|$ diffusion events per compartments (where z is the assumed constant number of neighbors of each compartment, ignoring boundary compartments), is $|\Omega|(r + z|L|)$. This augmented set of species and reactions is then simulated using the Gillespie algorithm in which the reaction propensities are appropriately scaled according to the compartment volumes.[4]

Crucial to the justification of this strategy to simulate a RDME is the assumption that, within each compartment, the reacting species can be considered homogeneously distributed so that one may use mass-action kinetics to determine the propensities. This assumption amounts to the condition that the timescale describing diffusion within compartments, τ_D, is much less than the timescale of chemical reactions, τ_C,

$$\tau_D \ll \tau_C. \qquad (34)$$

This comparison should be done for each diffusing and reacting species, and the timescale of the fastest reaction (taken as the inverse of the propensity) for each species should be used. If the condition holds, then the process of intra-compartmental diffusion will homogenize the solution faster than chemical reactions that occur, so the assumption of mass-action kinetics holds. Let the dimension of the space be d, the length of the (here assumed cubical) compartments be h, and the diffusion constant of a species be D. Then,

$$\tau_D \approx \frac{h^2}{2dD}. \qquad (35)$$

The Kuramoto length is given by

$$l_K = \sqrt{2dD\tau_C}, \qquad (36)$$

so one can see that the condition $\tau_D \ll \tau_C$ is equivalent to the condition $l_K \gg h$, and thus, one can enforce this condition by choosing a smaller compartment size h. For fixed total volume, there is a trade-off between h and $|\Omega|$, which determines the size of the augmented system and therefore the computational efficiency. The timescale of intra-compartment diffusion τ_D is approximately equal to the timescale of inter-compartment diffusion (which can be given as the inverse of $k_D = \frac{D}{h^2}$), so a third way of describing this condition is that the frequency of jumps between adjacent compartments is much greater than the frequency of chemical reactions inside the compartments, which can be checked empirically in simulation.[1,4] We note that, in the literature, these expressions may differ up to a constant coefficient depending on the reference.

In order to approximately describe the thermodynamics of this system, we distinguish between the inert solvent and the dilute, chemically reactive solutes. We assume here that the system is impermeable to the flow of either kind of species to the exterior. Within the system, all species are permeable (though local diffusion constants may be incorporated for each species[1]). We treat the solutes explicitly (at the level of their compartment copy numbers), whereas we model the solvent implicitly through an appropriate limit as in the steps leading to Eq. (21). To this end, we assume the existence of a laboratory timescale τ_l whose purpose is to define the temporal resolution of our measurements such that any changes occurring on a timescale longer than τ_l will be measured. We assume that τ_l is much longer than the timescale describing the local compartment fluctuations in the solvent copy numbers τ_s, yet shorter than but on the order of the timescale describing diffusion of the solutes τ_D. On this timescale τ_l, then in the time between the chemical reactions and diffusion events involving the solutes, the system is quasi-equilibrated with respect to fluctuations involving the fast process of solvent diffusion, and thus, the system may be assigned well-defined values of Gibbs free energy.[44] This hierarchy of timescales can be written as

$$\tau_s \ll \tau_l \lesssim \tau_D \ll \tau_C. \tag{37}$$

Typical ratios of the diffusion constants for solute to solvent are in the range of 1/10 to 1/100, placing τ_D/τ_s in the range of 10–100.[45]

On the timescale τ_l, there is enough temporal resolution to track the diffusion and chemical reactions of the solutes while allowing averaging over the fluctuations of the solvent. To describe activity occurring over the large grid of compartments for extended systems, we introduce new timescales τ_x^g, where x refers to any of the timescales defined above. If we hold the compartment size h and the chemical concentrations fixed and add more compartments to our system, the rates of solute diffusion events and chemical reactions occurring anywhere in the system will scale as $|\Omega|$, the number of compartments, and thus, the timescales needed to describe them scale as $|\Omega|^{-1}$, i.e., $\tau_l^g \sim \tau_l/|\Omega|$. One might be concerned that τ_l^g will be less than τ_s for large systems, ostensibly violating our requirement that we will be able to average over the solvent fluctuations to define quasi-equilibrated states. However, the timescale of solvent fluctuations across the whole grid, τ_s^g, will also scale inversely with $|\Omega|$, so the condition for being able to average over solvent fluctuations expressed in Eq. (37) does not ultimately depend on the number of compartments. In other words, for systems with many compartments, as long as Eq. (37) holds for one compartment, we can be sure that our laboratory timescale that describes the whole grid, τ_l^g, will be short enough to describe processes involving the solutes, while long enough to allow averaging over the fluctuations of the solvent occurring locally in each compartment.

We assume that the exterior of the system acts as a reservoir for the thermodynamic variables p and T. The volume V of the system also remains constant; however, under the assumption of the solvent being an incompressible liquid, the change in quantity pV is approximately zero, and for each reaction, the change in Gibbs free energy is equal to that of the Helmholtz free energy. Thus, it is inconsequential whether we consider p or V to be reservoir variables, and

we choose p in order to speak of the Gibbs free energy of the system. We can write the Gibbs free energy of the system as $G(\mathbf{N}, p, T)$, where $\mathbf{N} = \{\{N_{i,A}\}_{i \in M}\}_{A \in \Omega}$ represents the set of copy numbers of all solute and solvent species in each compartment in the grid. We further assume the compartments to be only weakly interacting, that is, they can exchange energy and particles, but the interaction of the two subsystems does not contribute a term to the Gibbs free energy of the system. This is equivalent to assuming that the Gibbs free energy of the compartments G_A is linearly additive,

$$G = \sum_{A \in \Omega} G_A, \tag{38}$$

without any terms of the form G_{AB}. To justify this, we first note that the interaction free energy between two adjacent compartments will primarily be due to the interaction of the solvent at the interface. This interfacial free energy will, even for mesoscopically sized compartments, be small compared to the free energy due to the bulk of the compartment. Furthermore, our main interest will be in changes in the Gibbs free energy of the system due to solute diffusion between compartments and chemical reactions, neither of which will significantly affect the interfacial free energy, so all terms of the form G_{AB} will drop out of the expression for ΔG.

In our approach of averaging over fluctuations in the solvent amounts and taking the limit that this average is large compared to the copy numbers of the solutes, we are choosing to neglect the changes in Gibbs free energy of the system owing to the solvent fluctuations. We justify this with an argument that these fluctuations are small compared to those resulting from the activity of the solute molecules; it also arises out of necessity due to the computational expense of tracking the solvent fluctuations. On the laboratory timescale τ_l, each compartment has an average copy number of solvent molecules, $\overline{N_{s,A}}$. The fluctuations in this quantity will have a standard deviation on the order of $\overline{N_{s,A}}^{1/2}$.[44] As indicated previously [to arrive at Eq. (21)], we avoid specifying $\overline{N_{s,A}}$ by taking the limit that it is much larger than the number of solute species, and thus, fluctuations in this quantity do not matter when computing nearly exact changes in the Gibbs free energy accompanying reactions involving the solutes. We need to establish that the change in Gibbs free energy of a compartment due to a fluctuation in the solvent copy number on the order of $\overline{N_{s,A}}^{1/2}$ is small compared to the change in Gibbs free energy accompanying chemical reactions and inter-compartment solute diffusion events and, thus, that the fluctuations in this quantity are not outweighing the changes that we are measuring. In Appendix C, we show that if the concentrations of solutes are very different in the two compartments, then this Gibbs free energy change is on the order of $\varepsilon \overline{N_{s,A}}^{1/2} k_B T$, where ε is the ratio of solutes to solvent. We show that this quantity is typically much less in magnitude than the change in Gibbs free energy accompanying a solute diffusion event. If the concentrations are nearly equal, then the Gibbs free energy is on the order of $\varepsilon k_B T$ and is thus negligibly small. We make the modeling choice to ignore these changes in the Gibbs free energy resulting from solvent fluctuations because they tend to be small and they do not represent the processes we are interested in, which involve the activity of the solute molecules.

To summarize, we track the changes in the Gibbs free energy of the system by computing a value of ΔG whenever a chemical

reaction occurs within a compartment or a solute diffusion event occurs between adjacent compartments. The timescales on which these events occur are much slower than that of solvent equilibration, allowing for well-defined values of G between the events. By the linear additivity of the compartments' free energies, any change in the free energy of a single compartment is equal to the change in free energy of the whole system (i.e., $\Delta G_A = \Delta G$). We assume that the diffusion of the solvent only contributes small, fluctuating, unbiased changes to the free energy that we ignore; we also assume that the amount of solvent is so large that one can take a limit and neglect the fluctuations in this quantity when computing the changes in free energy for processes involving the solute copy numbers.

C. ΔG of diffusion

To describe the change in Gibbs free energy accompanying a solute diffusion event between neighboring compartments, we use a similar approach to the one used above for chemical reactions. The key difference here is that, as opposed to multiple species being involved in a reaction taking place in a single compartment, we now have a single species involved in a diffusion event taking place between two compartments.

Consider species i diffusing from compartment A, where its initial copy number is $N_{i,A}$, to compartment B, where its initial copy number is $N_{i,B}$. The total copy numbers of molecules in compartments A and B are N_A and N_B, respectively. Assume there is just one spectator species constituting the solvent, labeled s with copy numbers $N_{s,A}$ and $N_{s,B}$ (we remove our uncertainty in the exact values of these numbers by taking a limit later). As a result of the diffusion event, we have the following changes in these quantities: $N_{i,A} \to N_{i,A} - 1$, $N_{i,B} \to N_{i,B} + 1$, $N_{s,A} \to N_{s,A}$, $N_{s,B} \to N_{s,B}$, $N_A \to N_A - 1$, and $N_B \to N_B + 1$. The initial Gibbs free energy is

$$G^{\text{initial}} = N_{i,A}\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,A}}{N_A}\right) + N_{i,B}\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,B}}{N_B}\right)$$
$$+ N_{s,A}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,A}}{N_A}\right) + N_{s,B}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,B}}{N_B}\right), \tag{39}$$

and the final Gibbs free energy is

$$G^{\text{final}} = (N_{i,A} - 1)\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,A} - 1}{N_A - 1}\right)$$
$$+ (N_{i,B} + 1)\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,B} + 1}{N_B + 1}\right)$$
$$+ N_{s,A}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,A}}{N_A - 1}\right)$$
$$+ N_{s,B}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,B}}{N_B + 1}\right). \tag{40}$$

The difference of these two expressions leads to an exact formula, similar to Eq. (16), which we omit here. Analogously to the approximation $N \gg \sigma$ made in the context of chemical reactions, here we assume that $N_A, N_B \gg 1$, which amounts to setting $N_A - 1 \approx N_A$ and $N_B + 1 \approx N_B$ in Eqs. (39) and (40). Using this approximation, we can express the change in Gibbs free energy as

$$\Delta G = k_B T \log \frac{(N_{i,A} - 1)^{(N_{i,A}-1)}}{N_{i,A}^{N_{i,A}}} \frac{(N_{i,B} + 1)^{(N_{i,B}+1)}}{N_{i,B}^{N_{i,B}}}. \tag{41}$$

We see that, analogously to making the approximation $N_i \gg \nu_i$ above, if we here take $N_{i,A}, N_{i,B} \gg 1$, then Eq. (41) reduces to the common expression

$$\Delta G = k_B T \log \frac{N_{i,B}}{N_{i,A}}. \tag{42}$$

The right-hand side of Eq. (42) will always be less than that of Eq. (41), which, although the difference is typically very slight, can lead to systematically biased calculations, as we describe in Sec. III.

The considerations leading to the main results of this paper [Eqs. (14), (21), and (41)] do not depend essentially on the assumption that the compartments comprising the simulation volume form a regular Cartesian grid (as illustrated in Fig. 1). A variety of strategies have been developed to discretize the simulation volume in a more sophisticated manner. These include unstructured meshes that use complex polygonal compartments useful for modeling curved surfaces, as well as adaptive meshes that use compartments with time-dependent boundaries that can increase computational efficiency.[46–48] Hybrid, multiscale methods also exist, which combine a mesoscopic compartment-based description of part of the simulation volume with either a macroscopic or microscopic description elsewhere in the volume.[49,50] Determining reaction and diffusion propensities in these frameworks is a somewhat complicated issue that we do not address here. Without undertaking a detailed analysis of how the above derivations of expressions for ΔG should be extended to apply in each of these computational frameworks, we observe that, given a method for calculating propensities leading to reaction and diffusion events, the expressions derived here can be applied directly to the case of an irregular grid of compartments, which encompasses several applications of interest. In an irregular grid, each compartment A has an arbitrary shape and volume V_A, with $V_A \neq V_B$, in general, for any pair A, B. It is important to ensure that for each of these compartments the assumptions regarding timescales described above still hold. If they do hold, then to apply the above results, it is only necessary to use compartment A's local value of $\Theta_A = N_{\text{Av}} V_A$ in place of Θ where it appears in the expressions for $\Delta G_{(1)}$, $\Delta G_{(2)}$, and $\Delta G_{(3)}$ [Eqs. (21), (25), and (28), respectively].

III. RESULTS

Here, we perform stochastic simulations to illustrate the effects of approximating ΔG for reaction and diffusion events when copy numbers are considered small and discrete. We use MEDYAN, a simulation platform designed to study active networks at high resolution, which is equipped with a RDME simulation engine as described above.[11] With this, we report on two different simulation setups to test the discrepancy between the nearly exact and approximate formulas for ΔG corresponding to reactions and to diffusion. We compare the nearly exact equations (21) and (41) for reactions and diffusion, respectively, with their approximate counterparts [Eqs. (28) and (42)]. We observe that only the nearly exact expressions result in sensible behavior, i.e., the rate of change of Gibbs free energy on average is 0 $k_B T$/s when the system is at equilibrium.

A. ΔG of reactions

To test the effect of approximation for ΔG of reactions, we consider the simple reaction scheme

$$A + B \overset{k_+}{\underset{k_-}{\rightleftharpoons}} C, \qquad (43)$$

where the rate constants to the right and left are $k_+ = 0.05 \, \mu M^{-1} \, s^{-1}$ and $k_- = 0.01 \, s^{-1}$, respectively, giving an equilibrium constant of $K_{eq} = k_-/k_+ = 0.2 \, \mu M$. We perform stochastic simulations with the Next Reaction Method in MEDYAN[51] using a single compartment of size $0.125 \, \mu m^3$ (i.e., in this example, there is no diffusion). To employ Eq. (21) in simulation, we first compute $\Delta G^0 - \sigma k_B T \log \Theta - \sigma k_B T = 3.71 \, k_B T$ for the forward reaction and $-3.71 \, k_B T$ for the reverse reaction (using $\Delta G^0 = k_B T \log K_{eq}$), and then, when each reaction fires during simulation, the quantity $\widetilde{Q}_{(1)}$ is computed from the instantaneous values of the copy numbers to determine $\Delta G_{(1)}$. A similar approach

is taken to employ Eq. (28). We begin with $N_A = 100$, $N_B = N_C = 50$ and repeat a simulation of 100 s duration 3000 times to obtain averages of the trajectory of rates $\partial_t \Delta G(t)$ resulting from the forward and reverse reactions [the notation $\partial_t \Delta G(t)$ implies that the measured rates vary smoothly as a function of time; however, this quantity is calculated as the total change in Gibbs free energy resulting from discretely timed chemical events during 1 s-long windows]. Figure 2 displays the results of these simulations. Note how, while the two trajectories bear close similarity, the equilibrium value of $\partial_t \Delta G(t)$ centers around 0 $k_B T/s$ for the nearly exact formulation [Eq. (21)], yet erroneously centers around $\sim -0.08 \, k_B T/s$ for the approximate version [Eq. (28)].

B. ΔG of diffusion

To test the effect of approximating ΔG for diffusion in a compartment-based reaction–diffusion scheme, we next employed MEDYAN to simulate diffusion of 1000 solute molecules with diffusion constant $20 \, \mu m^2 \, s^{-1}$ in a $2 \times 2 \times 2$ grid of compartments, each a
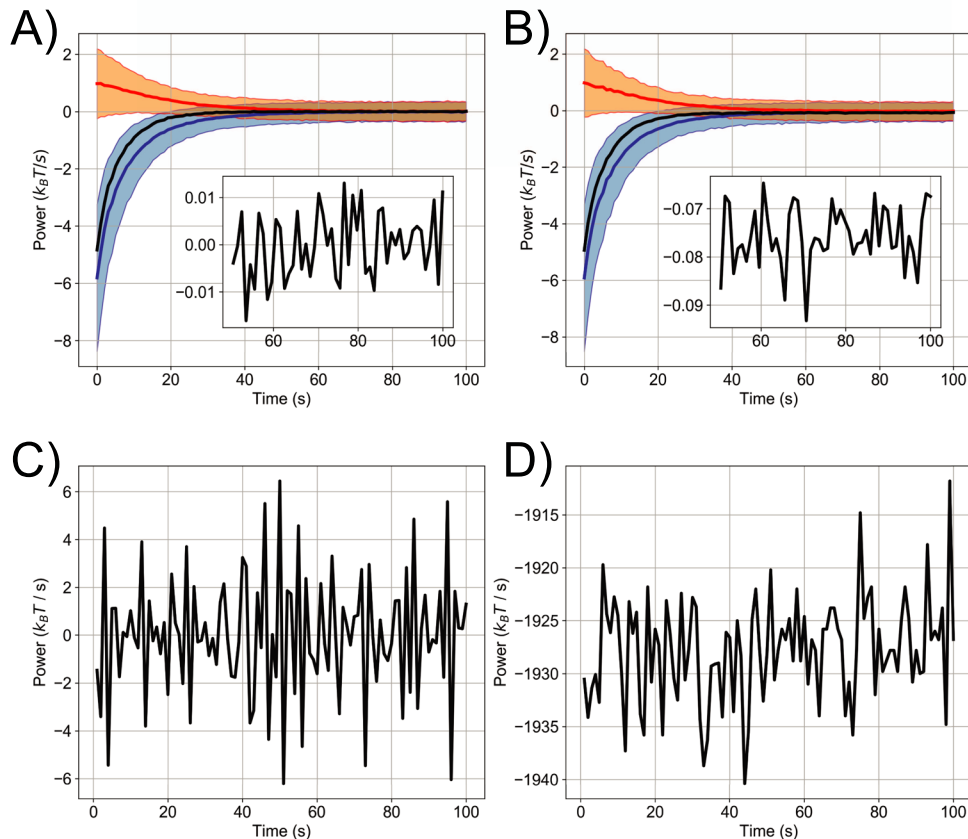
FIG. 2. Numerical results illustrating the difference between nearly exact and approximate formulations of ΔG for reactions and diffusion. (a) Averages over 3000 simulations of the chemical scheme $A + B \rightleftharpoons C$, using the nearly exact Eq. (21). The blue curve represents the trajectory of $\partial_t \Delta G$ resulting from the forward reaction, the red curve represents that from the reverse reaction, and the black curve represents their sum. Shaded regions represent the standard deviation over the 3000 repeated trials. The inset displays a blow-up of the black curve once the system has reached close to equilibrium. (b) The same as just described, but using the approximate Eq. (28). (c) A single trajectory of diffusion of 1000 molecules over a $1 \, \mu m^3$ cubic grid of eight compartments, beginning from a random initial spatial distribution. Values of ΔG are calculated using Eq. (41). (d) The same as just described; however, the values of ΔG are calculated using Eq. (42).

cube with volume 0.125 μm^3. The initial distribution of molecules is uniformly random over the compartment grid, and thus, the system begins near equilibrium and is then allowed to stochastically evolve for 100 s, i.e., the molecules hop randomly between adjacent compartments. For each diffusion event, the value of ΔG is determined using the nearly exact equation (28) for one run, and in another run, the approximate equation (42) is used. The difference in the trajectories of $\partial_t \Delta G$ for these simulations is stark, as displayed in Fig. 2. While the trajectory centers around 0 $k_B T$/s for the nearly exact formulation of ΔG, it erroneously centers around ~ -1925 $k_B T$/s for the approximate version. Diffusion events are very frequent in this system, occurring around 240 000 times/s (this number can be calculated from the parameters of the system and is also observed during simulations). Thus, since Eq. (42) is always less than the nearly exact quantity, even by a small amount on the order 0.05 $k_B T$ for this system, this systematic bias is amplified by the frequency of diffusion events to produce significant differences from the expected behavior, necessitating the use of a more exact formula for ΔG. Finally, we performed a simulation involving both reactions and diffusion across multiple compartments. We found again that only when the nearly exact formulas were used did the rate of change of Gibbs free energy center on 0 $k_B T$/s at equilibrium. It is not additionally illuminating to show the data, so we do not display it here.

IV. DISCUSSION

We have argued that when the copy numbers of the reactants and products are treated as small, discrete quantities, then certain approximations leading to the textbook formulas for ΔG of reaction and diffusion [Eqs. (2) and (3)] break down and lead to biased results. We emphasize that this is true only when the copy numbers and reaction occurrences are treated as discrete; when they are treated as continuous, one should use the textbook formulas. This can be shown by considering a continuous version of the chemical system described by Eq. (43). The time evolution of the concentrations of the chemical species is obtained by solving a system of ordinary differential equations that employ mass-action kinetics,

and from this solution, the rates $\partial_t \Delta G(t)$ resulting from the forward and reverse reactions are computed using the instantaneous values of the species' copy numbers that enter into Eqs. (21) and (28). Figure 3 displays the results of these calculations. Here, the total rate of $\partial_t \Delta G(t)$ only approaches 0 $k_B T$/s, as it must at equilibrium, when Eq. (28) is used.

When the copy numbers of the chemical species are treated as continuous, there is no notion of a single occurrence of a reaction; instead, the evolution of the system is parameterized by the continuous variable ξ that quantifies the extent of advancement of the reaction.[52] In this framework, which is adopted in classical thermodynamics, the copy number of any species never jumps instantaneously from N_i to $N_i + \nu_i$, and thus, the premise of the derivation presented above leading to Eq. (16) falls apart. This explains why the seemingly more accurate equation (21) is wrong when applied to chemical dynamics that are described using continuous variables to represent the copy number of species. When the chemical dynamics are modeled this way, the textbook expression [Eq. (28)] for the change in Gibbs free energy is valid. Another way to understand the difference in these expressions is to view Eq. (28) as giving the instantaneous slope of G with respect to the degree of advancement of the reaction when these quantities are viewed as continuous variables,[38,42] whereas Eq. (16) effectively integrates the continuously varying quantity dG through a single discrete reaction occurrence.

If the copy numbers of chemical species are not large compared to the stoichiometric coefficients (i.e., N_i is not much greater than ν_i), it is best to describe the chemical dynamics by treating the copy numbers as discrete variables participating in stochastically timed chemical reactions. This is the philosophy adopted by several recent models of intracellular environments, where copy numbers of molecules of interest are sometimes quite small. In these cases, adoption of the more exact expression for ΔG can not only improve precision but also ensure correct behavior. If concentration gradients are expected to be a strong source of entropy production (e.g., diffusion of monomeric actin along the lengths of filopodia, see Refs. 53–55), then using the nearly exact formulas presented here can ensure
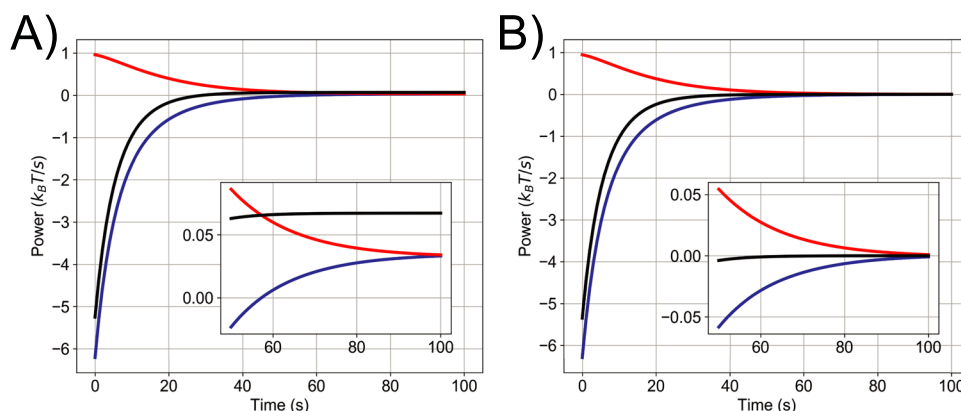
FIG. 3. Analytical results illustrating that, when copy numbers are treated as continuous variables, the textbook formula for ΔG of reaction should be used. (a) Calculation of the rates $\partial_t \Delta G(t)$ using Eq. (21) resulting from the forward (blue curve) and reverse (red curve) reactions, as well as their sum (black curve), as described in the main text. The inset shows a blow-up of these curves as the system approaches equilibrium. (b) The same as just described, but using Eq. (28).

that the resulting measurements of dissipation are not strongly biased.

ACKNOWLEDGMENTS

APPENDIX A: DISCRETE VARIABLES AND THE RELATION TO THE GIBBS–DUHEM EQUATION

The Gibbs–Duhem equation from classical thermodynamics states

$$\sum_{i \in M} N_i d\mu_i = -S dT + V dp = 0, \qquad (A1)$$

where M represents all chemical species in the system including the solvent, and the last equality holds at constant temperature and pressure.[27] Applying the product rule[56] to the expression for the Gibbs free energy, $G = \sum_{i \in M} N_i \mu_i$, we get

$$dG = \sum_{i \in M} \mu_i dN_i + \sum_{i \in M} N_i d\mu_i = \sum_{i \in M} \mu_i dN_i. \qquad (A2)$$

Evaluating $\sum_{i \in M} \mu_i dN_i$ using stoichiometric coefficients in place of dN_i can be shown to lead to the expression $\Delta G_{(3)}$.

If the copy numbers are small, we are better served using the discrete difference operator Δ and not the differential difference operator d. The product rule for the discrete difference operator applied to G gives

$$\Delta G = \sum_{i \in M} \mu_i \Delta N_i + \sum_{i \in M} N_i \Delta \mu_i + \sum_{i \in M} \Delta \mu_i \Delta N_i. \qquad (A3)$$

Here, we do not neglect the cross-term as we do in the differential product rule [Eq. (A2)]. We evaluate this expression term by term,

$$\sum_{i \in M} \mu_i \Delta N_i = \Delta G_{(3)}, \qquad (A4)$$

$$\sum_{i \in M} N_i \Delta \mu_i = k_B T \log \prod_{i \in R} \left(\frac{N_i - v_i}{N_i}\right)^{N_i} \prod_{j \in P} \left(\frac{N_j + v_j}{N_j}\right)^{N_j} \left(\frac{N}{N + \sigma}\right)^N, \qquad (A5)$$

$$\sum_{i \in M} \Delta N_i \Delta \mu_i = k_B T \log \prod_{i \in R} \left(\frac{N_i - v_i}{N_i}\right)^{-v_i} \prod_{j \in P} \left(\frac{N_j + v_j}{N_j}\right)^{v_j} \left(\frac{N}{N + \sigma}\right)^\sigma. \qquad (A6)$$

Comparing these to the expressions for the accuracy of the various approximations given in Appendix B, we have

$$\sum_{i \in M} N_i \Delta \mu_i + \sum_{i \in M} \Delta N_i \Delta \mu_i = \Delta G_{(0)} - \Delta G_{(3)}, \qquad (A7)$$

and thus, combining all the terms in Eq. (A3), we recover our exact expression $\Delta G_{(0)}$, and this approach can be seen as an alternate derivation of Eq. (14). To summarize, for small, discrete copy numbers, we obtain corrections to the expression for the change in Gibbs free energy accompanying chemical reactions that are not captured by the constraints imposed by the Gibbs–Duhem equation, which assumes that the copy numbers are continuous quantities.

APPENDIX B: ACCURACY OF THE APPROXIMATIONS

We calculate the accuracy of the approximations $\Delta G_{(1)}$, $\Delta G_{(2)}$, and $\Delta G_{(3)}$ by taking the difference of these quantities with $\Delta G_{(0)}$. We have

$$(\Delta G_{(1)} - \Delta G_{(0)})/k_B T = \log\left(\frac{N + \sigma}{N}\right)^{N+\sigma} - \sigma, \qquad (B1)$$

$$(\Delta G_{(2)} - \Delta G_{(0)})/k_B T = \log\left(\frac{N + \sigma}{N}\right)^{N+\sigma} + \log \prod_{i \in R}\left(\frac{N_i}{N_i - v_i}\right)^{N_i}$$
$$\times \prod_{j \in P}\left(\frac{N_j}{N_j + v_j}\right)^{N_j}, \qquad (B2)$$

$$(\Delta G_{(3)} - \Delta G_{(0)})/k_B T = \log\left(\frac{N + \sigma}{N}\right)^{N+\sigma} + \log \prod_{i \in R}\left(\frac{N_i}{N_i - v_i}\right)^{N_i - v_i}$$
$$\times \prod_{j \in P}\left(\frac{N_j}{N_j + v_j}\right)^{N_j + v_j}. \qquad (B3)$$

The accuracy of $\Delta G_{(1)}$ depends only on N for a given reaction, whereas the remaining approximations depend also on the values of N_i and N_j. These observations reflect the fact that, in order to arrive at the expression for $\Delta G_{(1)}$, we leveraged the size of N compared to σ, and to arrive at the expressions for $\Delta G_{(2)}$ and $\Delta G_{(3)}$, we further successively leveraged the sizes of N_i, N_j compared to v_i, v_j.

APPENDIX C: ΔG OF SOLVENT FLUCTUATIONS

Here, we first calculate an approximate expression for the change in Gibbs free energy of the system accompanying a fluctuation of n solvent molecules from compartment A to compartment B. We have

$$G^{\text{initial}} = \sum_{i \in L} N_{i,A}\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,A}}{N_A}\right) + N_{s,A}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,A}}{N_A}\right)$$
$$+ \sum_{i \in L} N_{i,B}\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,B}}{N_B}\right) + N_{s,B}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,B}}{N_B}\right), \qquad (C1)$$

where L is the set of solute species. After the transfer of n solvent molecules from A to B, we have

$$G^{\text{final}} = \sum_{i \in L} N_{i,A}\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,A}}{N_A - n}\right) + N_{s,A}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,A} - n}{N_A - n}\right)$$
$$+ \sum_{i \in L} N_{i,B}\left(\widetilde{\mu_i^0} + k_B T \log \frac{N_{i,B}}{N_B + n}\right) + N_{s,B}\left(\widetilde{\mu_s^{0,*}} + k_B T \log \frac{N_{s,B} + n}{N_B + n}\right). \qquad (C2)$$

Taking the difference and simplifying, we arrive at the exact expression

$$\Delta G = k_B T \log \frac{N_A^{N_A}}{(N_A - n)^{N_A - n}} \frac{(N_{s,A} - n)^{N_{s,A} - n}}{N_{s,A}^{N_{s,A}}} \frac{N_B^{N_B}}{(N_A + n)^{N_A + n}}$$
$$\times \frac{(N_{s,B} + n)^{N_{s,B} + n}}{N_{s,B}^{N_{s,B}}}. \qquad (C3)$$

To understand the magnitude of ΔG for typical values of n, $N_{s,A}$, and $N_{s,B}$ compared to N_A and N_B, we first make the assumption that the two compartments initially have the same number of solvent molecules, i.e., $N_{s,A} = N_{s,B} \equiv N_s$. Next, we assume that the solvent molecules dominate the proportion of total molecules, allowing us to write $N_A \approx N_B \equiv N$. These approximations will hold in the limit that the number of solute molecules is much less than the number of solvent molecules for each compartment. We next introduce the small parameters

$$\varepsilon_A = \frac{\sum_{i \in L} N_{i,A}}{N} \tag{C4}$$

and

$$\varepsilon_B = \frac{\sum_{i \in L} N_{i,B}}{N}, \tag{C5}$$

which capture the dilutions of the two compartments, and

$$\xi = \frac{n}{N}, \tag{C6}$$

which represents the relative size of the fluctuation. For $N = 10^9$, we typically have $\xi \sim 10^{-4.5}$ (since $n \sim N^{1/2}$) and $\varepsilon \sim 10^{-6}$. Substituting these parameters into Eq. (C3), we have

$$\Delta G = k_B T \log \frac{N^{2N}}{(N(1-\xi))^{N(1-\xi)}(N(1+\xi))^{N(1+\xi)}}$$
$$\times \frac{(N(1-\varepsilon_A-\xi))^{N(1-\varepsilon_A-\xi)}(N(1-\varepsilon_B+\xi))^{N(1-\varepsilon_B+\xi)}}{(N(1-\varepsilon_A))^{N(1-\varepsilon_A)}(N(1-\varepsilon_B))^{N(1-\varepsilon_B)}}. \tag{C7}$$

We expand this expression to first order in ε_A and ε_B and then to second order in ξ. The result is

$$\Delta G \approx (\varepsilon_A - \varepsilon_B)N\xi k_B T + \frac{1}{2}(\varepsilon_A + \varepsilon_B)N\xi^2 k_B T$$
$$= (\varepsilon_A - \varepsilon_B)n k_B T + \frac{1}{2}(\varepsilon_A + \varepsilon_B)\frac{n^2}{N} k_B T. \tag{C8}$$

The observed numerical agreement between Eqs. (C3) and (C8) is close for realistic values of the parameters: for $N_A = 10^9$, $N_B = 5 \times 10^8$, $\varepsilon_A = 10^{-6}$, $\varepsilon_B = 3 \times 10^{-6}$, and $n = N_A^{1/2}$, the prediction of Eq. (C3) is $-0.0632401\ k_B T$ and the prediction of Eq. (C8) is $-0.0632436\ k_B T$. The first term in Eq. (C8) will dominate if the solute dilutions in the two compartments are very different from each other. In this case, we may compare the size of this change in Gibbs free energy to that accompanying the diffusion of a solute from compartment B to compartment A. This latter change in Gibbs free energy will be approximately $k_B T \log \frac{\varepsilon_A}{\varepsilon_B}$. If we now set $\varepsilon_A = a\varepsilon_B$, where a is of order 1 (typically, it will fall in the range [1/10, 10]), then ΔG for the solvent fluctuation will be $k_B T(a-1)\varepsilon_B n$ and ΔG for the solute diffusion will be $k_B T \log a$. The product $\varepsilon_B n$ will be typically on the order of $\sim 10^{-1.5}$, so one can see that for typical values of the parameters the change in Gibbs free energy from a solute diffusion event will be significantly greater in magnitude than that from a solvent fluctuation. If the dilutions are very similar, then $\varepsilon_A - \varepsilon_B \approx 0$, and the second term in Eq. (C8) dominates. This term is on the order of $(\varepsilon_A + \varepsilon_B)k_B T$ since $\frac{n^2}{N} \sim 1$. These changes in Gibbs free energy will typically be much smaller than those accompanying a chemical reaction or inter-compartment diffusion of the solute. Thus, we may neglect the activity of the solvent in tracking the Gibbs free energy of the system.

REFERENCES

[1] R. Grima and S. Schnell, "Modelling reaction kinetics inside cells," Essays Biochem. **45**, 41–56 (2008).

[2] N. Tanaka and G. A. Papoian, "Reverse-engineering of biochemical reaction networks from spatio-temporal correlations of fluorescence fluctuations," J. Theor. Biol. **264**(2), 490–500 (2010).

[3] F. Baras and M. Malek Mansour, "Reaction-diffusion master equation: A comparison with microscopic simulations," Phys. Rev. E **54**(6), 6139 (1996).

[4] D. Bernstein, "Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm," Phys. Rev. E **71**(4), 041103 (2005).

[5] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," J. Phys. Chem. **81**(25), 2340–2361 (1977).

[6] J. Schaff, C. C. Fink, B. Slepchenko, J. H. Carson, and L. M. Loew, "A general computational framework for modeling cellular structure and function," Biophys. J. **73**(3), 1135–1146 (1997).

[7] L. M. Loew and C. S. James, "The virtual cell: A software environment for computational cell biology," Trends Biotechnol. **19**(10), 401–406 (2001).

[8] E. Roberts, J. E. Stone, and Z. Luthey-Schulten, "Lattice microbes: High-performance stochastic simulation method for the reaction-diffusion master equation," J. Comput. Chem. **34**(3), 245–255 (2013).

[9] M. J. Hallock and Z. Luthey-Schulten, "Improving reaction kernel performance in lattice microbes: Particle-wise propensities and run-time generated code," in *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)* (IEEE, 2016), pp. 428–434.

[10] J. Hattne, D. Fange, and J. Elf, "Stochastic reaction-diffusion simulation with mesoRD," Bioinformatics **21**(12), 2923–2924 (2005).

[11] K. Popov, J. Komianos, and G. A. Papoian, "MEDYAN: Mechanochemical simulations of contraction and polarity alignment in actomyosin networks," PLoS Comput. Biol. **12**(4), e1004877 (2016).

[12] C. Floyd, G. A. Papoian, and C. Jarzynski, "Quantifying dissipation in actomyosin networks," Interface Focus **9**(3), 20180078 (2019).

[13] M. Ander, B. Pedro, B. Di Ventura, J. Ferkinghoff-Borg, M. A. F. M. Foglierini, C. Lemerle, I. Tomas-Oliveira, and L. Serrano, "Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: Analysis of simple networks," Syst. Biol. **1**(1), 129–138 (2004).

[14] B. Drawert, S. Engblom, and A. Hellander, "URDME: A modular framework for stochastic simulation of reaction-transport processes in complex geometries," BMC Syst. Biol. **6**(1), 76 (2012).

[15] S. S. Andrews, J. A. Nathan, R. Brent, and P. A. Adam, "Detailed simulations of cell biology with Smoldyn 2.1," PLoS Comput. Biol. **6**(3), e1000705 (2010).

[16] M. Vigelius, A. Lane, and B. Meyer, "Accelerating reaction–diffusion simulations with general-purpose graphics processing units," Bioinformatics **27**(2), 288–290 (2010).

[17] S. Wils and E. De Schutter, "Steps: Modeling and simulating complex reaction-diffusion systems with python," Front. Neuroinf. **3**, 15 (2009).

[18] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics* (Courier Corporation, 2004).

[19] T. M. Earnest, J. A. Cole, and Z. Luthey-Schulten, "Simulating biological processes: Stochastic physics from whole cells to colonies," Rep. Prog. Phys. **81**(5), 052601 (2018).

[20] D. A. Beard and H. Qian, "Relationship between thermodynamic driving force and one-way fluxes in reversible processes," PLoS One **2**(1), e144 (2007).

[21] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, 3rd ed. (Interscience, New York, 1967).

[22] R. Rao and M. Esposito, "Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics," Phys. Rev. X **6**(4), 041064 (2016).

[23] J. L. England, "Dissipative adaptation in driven self-assembly," Nat. Nanotechnol. **10**(11), 919 (2015).

[24] D. S. Seara, V. Yadav, I. Linsmeier, A. P. Tabatabai, W. O. Patrick, S. M. A. Tabei, S. Banerjee, and M. P. Murrell, "Entropy production rate is maximized in non-contractile actomyosin," Nat. Commun. **9**(1), 4948 (2018).

[25] J. L. England, "Statistical physics of self-replication," J. Chem. Phys. **139**(12), 121923 (2013).

23 September 2023 23:53:42

[26] I. Prigogine and G. Nicolis, "Biological order, structure and instabilities," Q. Rev. Biophys. **4**(2-3), 107–148 (1971).

[27] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, 1998).

[28] P. Attard, *Thermodynamics and Statistical Mechanics* (CUP Archive, 2002).

[29] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations* (Wiley, 1977), pp. 339–426.

[30] H. Lodish, J. E. Darnell, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira *et al.*, *Molecular Cell Biology* (Macmillan, 2008).

[31] R. Phillips, J. Theriot, J. Kondev, and H. Garcia, *Physical Biology of the Cell* (Garland Science, 2012).

[32] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell*, 4th ed. (Garland, 2002).

[33] P. W. Atkins, J. De Paula, and J. Keeler, *Atkins' Physical Chemistry* (Oxford University Press, 2018).

[34] D. A. Beard and H. Qian, *Chemical Biophysics: Quantitative Analysis of Cellular Systems* (Cambridge University Press, 2008).

[35] D. A. McQuarrie and J. D. Simon, *Physical Chemistry: A Molecular Approach* (University Science Books, Sausalito, CA, 1997), Vol. 1.

[36] T. L. Hill, "Perspective: Nanothermodynamics," Nano Lett. **1**, 111 (2001).

[37] T. L. Hill, *Thermodynamics of Small Systems* (Courier Corporation, 1994).

[38] J. de Heer, "The free energy charge accompanying a chemical reaction and the Gibbs-Duhem equation," J. Chem. Educ. **63**(11), 950 (1986).

[39] T. Engel, P. Reid *et al.*, *Physical Chemistry* (Pearson, 2006).

[40] The argument of the logarithm, $\frac{N_A}{\Theta_A}$, has dimensions of concentration, however it is understood that such quantities are taken with reference to some standard concentration, typically 1M.

[41] D. M. Anderson, J. D. Benson, and A. J. Kearsley, "Foundations of modeling in cryobiology—I: Concentration, Gibbs energy, and chemical potential relationships," Cryobiology **69**(3), 349–360 (2014).

[42] G. M. Barrow, "Free energy and equilibrium: The basis of G0 = -RT in K for reactions in solution," J. Chem. Educ. **60**(8), 648 (1983).

[43] We have $-\log\left(\frac{N+\sigma}{N}\right)^{(N+\sigma)} = -\log\left(1 + \frac{\sigma}{N}\right)^{(N+\sigma)} \approx -\sigma$ for large N.

[44] Y. Mishin, "Thermodynamic theory of equilibrium fluctuations," Ann. Phys. **363**, 48–97 (2015).

[45] R. Milo and R. Phillips, *Cell Biology by the Numbers* (Garland Science, 2015).

[46] S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt, "Simulation of stochastic reaction-diffusion processes on unstructured meshes," SIAM J. Sci. Comput. **31**(3), 1774–1797 (2009).

[47] S. A. Isaacson and C. S. Peskin, "Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations," SIAM J. Sci. Comput. **28**(1), 47–74 (2006).

[48] B. Bayati, P. Chatelain, and P. Koumoutsakos, "Adaptive mesh refinement for stochastic reaction–diffusion processes," J. Comput. Phys. **230**(1), 13–26 (2011).

[49] L. Ferm, A. Hellander, and P. Lötstedt, "An adaptive algorithm for simulation of stochastic reaction–diffusion processes," J. Comput. Phys. **229**(2), 343–360 (2010).

[50] M. B. Flegg, S. J. Chapman, and R. Erban, "The two-regime method for optimizing stochastic reaction–diffusion simulations," J. R. Soc., Interface **9**(70), 859–868 (2011).

[51] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," J. Phys. Chem. A **104**(9), 1876–1889 (2000).

[52] J. Borge, "Reviewing some crucial concepts of Gibbs energy in chemical equilibrium using a computer-assisted, guided-problem-solving approach," J. Chem. Educ. **92**(2), 296–304 (2014).

[53] R. Erban, M. B. Flegg, and G. A. Papoian, "Multiscale stochastic reaction–diffusion modeling: Application to actin dynamics in filopodia," Bull. Math. Biol. **76**(4), 799–818 (2014).

[54] D. Ulrich, G. A. Papoian, and R. Erban, "Steric effects induce geometric remodeling of actin bundles in filopodia," Biophys. J. **110**(9), 2066–2075 (2016).

[55] Y. Lan and G. A. Papoian, "The stochastic dynamics of filopodial growth," Biophys. J. **94**(10), 3839–3852 (2008).

[56] The product rule can be derived as $d(fg) = (f + df)(g + dg) - fg = fdg + gdf + dfdg = fdg + gdf$ since the term $dfdg$ is assumed to be negligible. We give this reminder to emphasize the presence of the cross term, which we do not neglect in the discrete case.

23 September 2023 23:53:42