Thermodynamic Uncertainty Relation for Biomolecular Processes
Andre C. Barato and Udo Seifert
II. Institut für Theoretische Physik, Universität Stuttgart, 70550 Stuttgart, Germany
(Received 29 October 2014; published 15 April 2015)

Biomolecular systems like molecular motors or pumps, transcription and translation machinery, and other enzymatic reactions, can be described as Markov processes on a suitable network. We show quite generally that, in a steady state, the dispersion of observables, like the number of consumed or produced molecules or the number of steps of a motor, is constrained by the thermodynamic cost of generating it. An uncertainty $\epsilon$ requires at least a cost of $2k_B T/\epsilon^2$ independent of the time required to generate the output.

Biomolecular processes are generally out of equilibrium and dissipative, with the associated free energy consumption coming most commonly from adenosine triphosphate (ATP) hydrolysis. The role of energy dissipation in a variety of processes related to biological information processing has received much attention recently [1–14], to give just one class of examples for which one tries to uncover fundamental limits involving energy dissipation in biomolecular systems.

Chemical reactions catalyzed by enzymes are of central importance for many cellular processes. Prominent examples are molecular motors [15–19], which convert chemical free energy from ATP into mechanical work. In this case, an observable of interest is the number of steps the motor made. Another commonly analyzed output in enzymatic kinetics is the number of product molecules generated by an enzymatic reaction, for which the Michaelis-Menten scheme provides a paradigmatic case [2].

Quite generally, chemical reactions are well described by stochastic processes. An observable, like the rate of consumed substrate molecules or the number of steps of a motor on a track, is a random variable subjected to thermal fluctuations. Single molecule experiments [20–24] provide detailed quantitative data on such random quantities. Obtaining information about the underlying chemical reaction scheme through the measurement of fluctuations constitutes a field called statistical kinetics [25–28]. A central result in this field is the fact that the Fano factor quantifying fluctuations provides a lower bound on the number of states involved in an enzymatic cycle [15,28].

For a nonzero mean output, the chemical potential difference (or affinity) driving an enzymatic reaction must also be nonzero, leading to a free energy cost. Is there a fundamental relation between the relative uncertainty associated with the observable quantifying the output and the free energy cost of sustaining the biomolecular process generating it?

In this Letter, we show that such a general bound does, indeed, exist. Specifically, for any process running for a time $t$, we show that the product of the average dissipated heat and the square of the relative uncertainty of a generic observable is independent of $t$ and bounded by $2k_B T$. This uncertainty relation is valid for general networks and can be proved within linear response theory. Beyond linear response theory, we show it analytically for unicyclic networks and verify it numerically for several multicyclic networks. As an illustration of a specific consequence of our results, we obtain a new bound on the Fano factor for unicyclic networks which depends on the affinity driving the process.

The observables we consider here arise from counting the number of transitions between states, which is different from a random variable counting the fraction of time spent in a state. For this latter quantity, the relative uncertainty is finite even in equilibrium, as is the case of a cell estimating the concentration of an external environment by counting the fraction of time receptors are bound [29,30]. The role of dissipation for this problem has been recently studied in [5,13].

Our main result can be introduced using, arguably, the simplest example of a nonequilibrium chemical reaction catalyzed by an enzyme, which is a biased random walk where a single step is interpreted as the completion of an enzymatic cycle. Steps to the right happen with a rate $k^+$, those to the left with a rate $k^-$. After a time $t$, on average, $\langle X \rangle = (k^+ - k^-)t$ steps have occurred, with the number of steps $X$ corresponding to the observable of interest. Specific realizations of this random process show a variance $\langle (X - \langle X \rangle)^2 \rangle = (k^+ + k^-)t$ [31]. The squared relative uncertainty of the observable is

$$\epsilon^2 \equiv \langle (X^2 - \langle X \rangle^2) \rangle / \langle X \rangle^2 = (k^+ + k^-) / [(k^+ - k^-)^2 t].$$

Assuming an external environment of fixed temperature $T$, the transition rates are given by the local detailed balance relation

$$k^+/k^- = \exp(A/k_B T).$$
where $\mathcal{A}$ is the affinity driving the process and $k_B$ Boltzmann’s constant. The thermodynamic cost of generating this output is given by the entropy production rate, which reads [32]

$$\sigma = (k^+ - k^-)A/T,$$

(3)

leading to a total dissipation after time $t$ of $T\sigma t$. In equilibrium, i.e., for $k^+ = k^-$, there is no dissipation and the uncertainty $\epsilon$ diverges.

The trade-off between precision and dissipation is captured by the crucial product $Q$ of total dissipation and the square of the relative uncertainty,

$$Q \equiv T\sigma t \epsilon^2 = A \coth[A/(2k_BT)] \geq 2k_BT,$$

(4)

where we used Eqs. (1), (2), and (3). This thermodynamic uncertainty relation shows that a more precise output requires a higher thermodynamic cost independent of the uncertainty relation shows that a more precise output

where this sum is over all affinities $A_\beta$, including the case $\beta = \alpha$, and $E_i$ is the equilibrium free energy of state $i$.

In the stationary state, the velocity (or probability current) and diffusion constant associated with $X_\alpha$ are defined as

$$J_\alpha \equiv \langle X_\alpha \rangle/t,$$

(6)

and

$$D_\alpha \equiv \langle (X_\alpha^2) - \langle X_\alpha \rangle^2 \rangle/(2t),$$

(7)

respectively. The squared relative uncertainty then reads

$$\epsilon_\alpha^2 \equiv \langle (X_\alpha^2) - \langle X_\alpha \rangle^2 \rangle/\langle X_\alpha \rangle^2 = 2D_\alpha/J_\alpha^2.$$

(8)

While the probability current has a simple form in terms of the stationary probability distribution, namely, $J_\alpha = \sum_{ij} d_{ij}^\alpha (P_{ij} - P_{ji})$, a general formula for the diffusion constant is more involved and will be discussed below. The entropy production rate is [32]

$$\sigma = \sum_\beta J_\beta A_\beta,$$

(9)

where $\beta$ runs over all affinities. For example, for a molecular motor this sum has two terms: one affinity is the chemical potential difference driving the motor with the rate of ATP consumption as the associated current, the other affinity is the mechanical force, and the respective current is the velocity of the motor. The (dimensionless) product (4) for a general network is then defined as

$$Q_\alpha \equiv \sigma t \epsilon_\alpha^2 = 2\sigma D_\alpha/J_\alpha^2.$$

(10)

Within linear response theory [32], which is valid close to equilibrium where the affinities $A_\beta$ are small, a current can be expressed by the affinities as

$$J_\beta = \sum_\gamma L_{\beta\gamma} A_\gamma,$$

(11)

where the Onsager coefficients are defined as

$$L_{\beta\gamma} \equiv \partial_\beta J_\gamma \Big|_{A=0} = L_{\gamma\beta}.$$

(12)

From Eqs. (9) and (11), the entropy production within linear response reads $\sigma = \sum_\beta J_\beta \epsilon_\beta^2 A_\beta$. Moreover, the diffusion constant is given by the Einstein relation $D_\alpha = L_{\alpha\alpha}$ [33], which from Eq. (8), leads to $\epsilon_\alpha^{-1}$ being linear in the affinities. Hence, Eq. (10) becomes...
\[ Q_\alpha = 2 \sum_{\varphi \neq \alpha} \frac{L_{\varphi \alpha} A_{\varphi} A_{\varphi}^T}{\sum_{\varphi \neq \alpha} G_{\varphi \alpha} A_{\varphi} A_{\varphi}} = 2 \left( \frac{1}{\langle I \rangle^2} \right) \] (13)

where \( G_{\varphi \alpha} = (L_{\varphi \alpha} - L_{\alpha \varphi} L_{\alpha \varphi}) \). Using the fact that the Onsager matrix \( L \) is positive semidefinite, it is possible to prove that \( G \) is also a positive semidefinite matrix [34]. Hence, we have established \( Q_\alpha \geq 2 \) within linear response theory. Note that equality is reached in the case of only one nonzero affinity, i.e., \( A_{\varphi} = 0 \) for \( \varphi \neq \alpha \).

In the calculations that follow, we use elegant expressions obtained by Koza [36,37] for the velocity and diffusion coefficient, which are valid for a general network of states. For these expressions, we need a modified generator associated with \( X_\alpha \), which is a \( N \)-dimensional square matrix with elements [38]

\[ [C'(z)]_{ij} = \begin{cases} k_{ij} \exp(zd_{ij}) & \text{if } i \neq j \\ -\sum_j k_{ij} & \text{if } i = j \end{cases}. \] (14)

A set of coefficients \( C_n(z) \) is defined through the characteristic polynomial of this matrix as

\[ \det[yI - L'(z)] = \sum_{n=0}^{N} C_n(z)y^n, \] (15)

where \( I \) represents the identity matrix. Using these coefficients, which are functions of the transition rates, the current and diffusion coefficient can be written as [36]

\[ J_\alpha = -C'_0/C_1, \] (16)

and

\[ D_\alpha = -(C'_0 + 2C'_1J_\alpha + 2C_2J_1^2)/(2C_1), \] (17)

where \( C_n \equiv C_n(0) \) and the primes denote derivatives with respect to \( z \) taken at \( z = 0 \). A full derivation for these expressions is given in [34].

We first consider an arbitrary unicyclic model with \( N \) states [39]. The transition rate from state \( n \) to state \( n + 1 \) \((n + 1)\) is denoted \( k_{n}^{+} \) \( (k_{n}^{-}) \), where \( n = 0, 1, \ldots, N - 1 \). The output \( X \) counts the number of completed cycles. It is sufficient to count the number of transitions through one of the links in the cycle, which we choose to be the link between states 0 and 1. The generalized distance associated with \( X \) is then \( d_{01} = -d_{10} = 1 \) and \( d_{ij} = 0 \) for \( i \neq j \).

The cycle affinity is

\[ \mathcal{A} = \ln(\Gamma^{+}/\Gamma^{-}), \] (18)

where \( \Gamma^{+} = \prod_{i=0}^{N-1} k_{i}^{+} \) and \( \Gamma^{-} = \prod_{i=0}^{N-1} k_{i}^{-} \). An example of such a unicyclic machine with \( N = 3 \) is an enzyme \( E \) that consumes ATP according to the scheme

\[ E + ATP \rightleftharpoons k_{1}^{+} ET \rightleftharpoons k_{2}^{-} ED \rightleftharpoons P_{i} \rightleftharpoons E + ADP + P_{i}, \] (19)

where ADP stands for adenosine diphosphate, \( P_{i} \) for phosphate, \( ET \) (\( ED \)) represents the enzyme with an ATP (ADP) bound to it. In this case, the variable \( X \) is the number of consumed ATP molecules and the affinity is given by the chemical potential difference \( \mathcal{A} = \mu_{\text{ATP}} - \mu_{\text{ADP}} - \mu_{P} \).

We can show that, for a given number of states \( N \) and affinity \( \mathcal{A} \), the product \( Q \) reaches its minimal value for uniform rates, i.e., \( k_{i}^{+} = (\Gamma^{+})^{1/N} \) and \( k_{i}^{-} = (\Gamma^{-})^{1/N} \) independent of \( i \), leading to the bound [34]

\[ Q \geq \langle A/N \rangle \coth[\mathcal{A}/(2N)] \geq 2, \] (20)

which is in agreement with Eq. (4) that corresponds to \( N = 1 \). The bound (20) gives the minimal dissipation required to realize an uncertainty \( \epsilon \) for given affinity \( \mathcal{A} \) and number of states \( N \). This bound is an increasing function of \( \mathcal{A} \); hence, \( Q \) is minimal for \( \mathcal{A} \to 0 \) where \( Q \to 2 \). A related quantity, defined as the ratio of a “barometric” force and an “Einstein” force, has been considered in [40], where a bound similar to \( Q \geq 2 \) has been shown to hold for the case \( N = 2 \) within a calculation keeping terms up to second order in the affinity \( \mathcal{A} \).

We now turn to a specific example showing how this new constraint involving fluctuations and energetic cost can be turned into a diagnostic tool to unveil a structural property of an enzymatic cycle. A quantity closely related to the relative uncertainty is the Fano factor

\[ F \equiv \langle X^2 \rangle - \langle X \rangle^2 / \langle X \rangle = 2D/J, \] (21)

which gives a measure of the dispersion associated with \( X \), where \( X \) counts the output of an enzymatic cycle. For unicyclic networks, this Fano factor is known to be bounded from below by \( 1/N [15,28] \). Measurements of the Fano factor can then be used to obtain a bound on the number of states of an underlying enzymatic cycle [28]. Our new bound (20) implies

\[ F = Q/\mathcal{A} \geq (1/N) \coth[\mathcal{A}/(2N)]. \] (22)

For a diverging affinity, which is the case in chemical reaction schemes where at least one backward transition rate is assumed to be zero, this bound becomes the known one \( F \geq 1/N \). For experiments where substrate and product concentrations are kept fixed and, consequently, the value of the affinity is known, as for example in [41], our stronger bound in Eq. (22) constrains, even further, the number of states in such an enzymatic cycle.

Let us turn again to multicyclic networks. Within the linear response regime, we have established above [after Eq. (13)] that \( Q_\alpha \) reaches the bound 2 for the case where all affinities but \( \mathcal{A}_\alpha \) are zero. For unicyclic networks, which is
the paradigmatic case for a system with one affinity, we have proved that the bound holds arbitrarily far from equilibrium, being reached only in the linear response regime. In order to provide full evidence that our main result \( Q_\alpha \geq 2 \) is, indeed, universal, we now analyze multicyclic networks beyond linear response. In this case, we have to take specific systems. As a first example, we consider a model with an enzyme \( E \) that can consume two different substrates \( S_1 \) and \( S_2 \) and generates product \( P \), see Fig. 1. Two enzymatic cycles of this model are

\[
E + S_1 \xrightleftharpoons[k_{31}]{k_{13}} ES_1 \xrightleftharpoons[k_{34}]{k_{41}} EP \xrightleftharpoons[k_{14}]{k_{42}} E + P,
\]

\[
E + S_2 \xrightleftharpoons[k_{31}]{k_{13}} ES_2 \xrightleftharpoons[k_{34}]{k_{41}} EP \xrightleftharpoons[k_{14}]{k_{42}} E + P,
\]

where the enzyme states are identified as \( E \cong 1 \), \( ES_1 \cong 2 \), \( ES_2 \cong 3 \), and \( EP \cong 4 \). The affinity of the cycle involving substrate \( S_1 \) (\( S_2 \)) is given by the chemical potential difference \( A_1 = \mu_1 - \mu_P \) (\( A_2 = \mu_2 - \mu_P \)). The relations between these affinities and the transition rates are \( A_1 = \ln[k_{12}k_{24}k_{41}k_{21}k_{13}] \) and \( A_2 = \ln[k_{13}k_{14}k_{34}k_{31}] \). There is also a third cycle \( 1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 1 \), in which an \( S_1 \) is consumed and an \( S_2 \) produced. Its affinity is not independent but rather given by \( A_1 - A_2 \).

The average rate at which \( S_1 \) molecules are consumed is given by the current \( J_1 \equiv P_1k_{12} - P_2k_{21} \), while the rate of \( S_2 \) consumption is \( J_2 \equiv P_1k_{13} - P_3k_{31} \). The entropy production can be conveniently written as a sum over terms which are the product of a cycle affinity and a current \[32\]. For the present model, it becomes

\[
\sigma = A_1J_1 + A_2J_2.
\]

As an observable of interest, we choose the number of consumed \( S_1 \) molecules, for which we can calculate the associated product \( Q_1 \) with formulas \( 10 \), \( 16 \), \( 17 \), and \( 24 \). The resulting function of the ten transition rates is too cumbersome to show. Both by minimizing this function numerically and by evaluating it for randomly chosen transition rates, we find that the uncertainty relation \( Q_1 \geq 2 \) is respected. In order to verify whether this result is particular to this network topology we have analyzed six other multicyclic networks, which do not share any particular symmetry \[34\]. For all these networks, our uncertainty relation \( Q_\alpha \geq 2 \) is fulfilled. In all cases numerical minimization of \( Q_\alpha \) leads to a minimum compatible with \( 2 \) that is reached in the linear response regime. Based on these results, we conjecture that \( Q_\alpha \geq 2 \) also for general multicyclic networks beyond linear response.

In conclusion, for nonequilibrium stationary states, we have conjectured the fundamental limit \( 2k_B T/\epsilon^2 \) on the minimal dissipation required to generate an output with small relative uncertainty \( \epsilon \). This bound can be saturated close to equilibrium with only one independent affinity driving the process. This uncertainty relation provides a universal link between the “precision” of a molecular machine and the cost of maintaining it. As one specific application, we have shown how a bound on the number of states involved in the enzymatic cycle related to the Fano factor can be improved provided the affinity is known. More broadly, one can expect similar signatures of our fundamental relation quantifying the minimal energetic cost for reaching small uncertainty, i.e., high precision, to show up in any biochemical or biophysical process at fixed temperature that can be described by a Markov network \[42\]. Exploring whether and how this balance between fluctuations and energetics has guided the evolution of chemical reaction networks in living systems constitutes one intriguing perspective of our approach. Our fundamental relationship between minimal dissipated heat and uncertainty is based, first, on exact results in the linear response regime. Second, we have proved it for unicyclic networks arbitrarily far from equilibrium. Third, for multicyclic networks far from equilibrium, we have numerical evidence for several different networks. We could not provide a formal proof for arbitrary networks, and we expect that the method used for unicyclic networks cannot be generalized to multicyclic networks, as it requires an expression for the diffusion coefficient in terms of the transition rates.

On the technical level, investigating possible generalizations of the affinity dependent bound on the Fano factor in Eq. \( 22 \) to multicyclic networks could lead to further new bounds in statistical kinetics. Likewise, it would be interesting to explore whether one can derive bounds involving higher order cumulants. Finally, we emphasize that an algebraic proof of the uncertainty relation in the multicyclic case beyond the linear response regime looks like a serious challenge.


Similarly, the generalization to machines involving coupling to two heat baths should be straightforward.