

Research



Cite this article: Marsland III R, Cui W, Horowitz JM. 2019 The thermodynamic uncertainty relation in biochemical oscillations. *J. R. Soc. Interface* **16**: 20190098. <http://dx.doi.org/10.1098/rsif.2019.0098>

Received: 15 February 2019

Accepted: 2 April 2019

Subject Category:

Life Sciences – Physics interface

Subject Areas:

biophysics

Keywords:

non-equilibrium statistical mechanics, stochastic thermodynamics, circadian oscillations, chemical reaction networks

Author for correspondence:

Robert Marsland III

e-mail: marsland@bu.edu

The thermodynamic uncertainty relation in biochemical oscillations

Robert Marsland III¹, Wenping Cui^{1,2} and Jordan M. Horowitz^{3,4,5}

¹Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215, USA

²Department of Physics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA

³Physics of Living Systems Group, Department of Physics, Massachusetts Institute of Technology, 400 Technology Square, Cambridge, MA 02139, USA

⁴Department of Biophysics, and ⁵Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA

RM, 0000-0002-5007-6877; JMH, 0000-0002-9139-0811

Living systems regulate many aspects of their behaviour through periodic oscillations of molecular concentrations, which function as ‘biochemical clocks.’ The chemical reactions that drive these clocks are intrinsically stochastic at the molecular level, so that the duration of a full oscillation cycle is subject to random fluctuations. Their success in carrying out their biological function is thought to depend on the degree to which these fluctuations in the cycle period can be suppressed. Biochemical oscillators also require a constant supply of free energy in order to break detailed balance and maintain their cyclic dynamics. For a given free energy budget, the recently discovered ‘thermodynamic uncertainty relation’ yields the magnitude of period fluctuations in the most precise conceivable free-running clock. In this paper, we show that computational models of real biochemical clocks severely underperform this optimum, with fluctuations several orders of magnitude larger than the theoretical minimum. We argue that this suboptimal performance is due to the small number of internal states per molecule in these models, combined with the high level of thermodynamic force required to maintain the system in the oscillatory phase. We introduce a new model with a tunable number of internal states per molecule and confirm that it approaches the optimal precision as this number increases.

1. Introduction

Many living systems regulate their behaviour using an internal ‘clock’, synchronized to the daily cycles of light and darkness. In the past 15 years, the isolation of the key components of several bacterial circadian clocks has opened the door to systematic and quantitative study of this phenomenon. In particular, a set of three proteins purified from the bacterium *Synechococcus elongatus* are capable of executing sustained periodic oscillations *in vitro* when supplied with ATP [1]. One of the proteins, KaiC, executes a cycle in a space of four possible phosphorylation states, as illustrated in figure 1. This cycle is coupled to the periodic association and dissociation from the other two proteins, KaiA and KaiB.

Steady oscillations break detailed balance and must be powered by a chemical potential gradient or other free energy source. In this system and in related experiments and simulations, it is commonly observed that the oscillator precision decreases as this thermodynamic driving force is reduced [3]. At the same time, recent theoretical work indicates that the precision of a generic biochemical clock is bounded from above by a number that also decreases with decreasing entropy production per cycle [4–11]. This has led to speculation that this universal bound may provide valuable information about the design principles behind real biochemical clocks.

So far, most discussion of this connection has focused on models with cyclic dynamics hard-wired into the dynamical rules [6,11,12]. But real biochemical oscillators operate in a high-dimensional state space of concentration profiles,

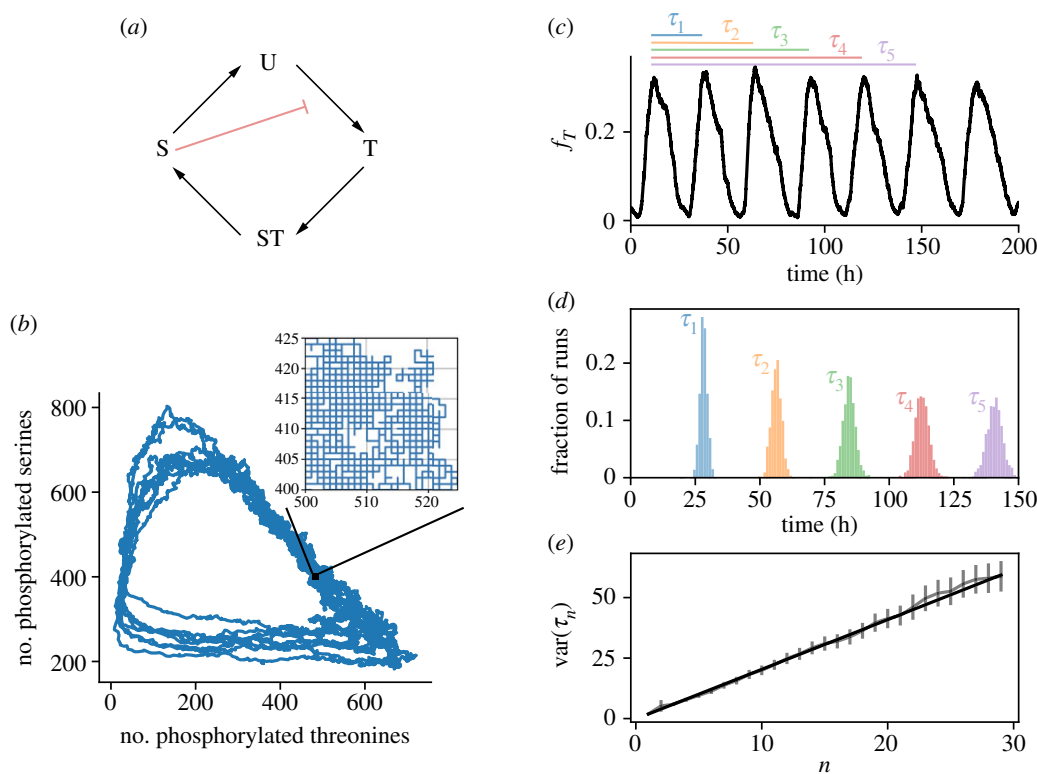


Figure 1. Coherent cycles in a biochemical oscillator. (a) Schematic of the KaiC biochemical oscillator. This simplified diagram shows four different internal states of the KaiC molecule, labelled U, T, S and ST (unphosphorylated, phosphorylated on threonine, phosphorylated on serine and phosphorylated on both residues). The molecules execute cycles around these four states in the indicated direction. They interact with each other via additional molecular components, in such a way that molecules in state S slow down the forward reaction rate for other molecules in state U. (b) Simulated trajectory from a detailed kinetic model of the KaiC system (adapted from [2]) with 360 interacting KaiC hexamers. The state space of this system is described by the list of copy numbers of all the molecular subspecies. Here, we have projected the state onto a two-dimensional plane, spanned by the copy numbers of monomers in states T and S. The inset gives a magnified view of a small portion of the plot, showing the discrete reaction steps caused by single phosphorylation and dephosphorylation events. (c) Time-evolution of the fraction f_T of monomers in state T. (d) Histograms of the time τ_n required for 1200 independent sets of 360 interacting KaiC hexamers to complete n collective cycles, for $n = 1$ through 10. (e) Variances $\text{var}(\tau_n)$ of the histograms as a function of n . Error bars are bootstrapped 95% confidence intervals. Black line is a linear fit, with slope $D = 2.05 \pm 0.05 \text{ h}^2$. (Online version in colour.)

and the cyclic behaviour is an emergent, collective phenomenon [3,6,13]. These oscillators typically exhibit a non-equilibrium phase transition at a finite value of entropy production per cycle. As this threshold is approached from above, the oscillations become more noisy due to critical fluctuations [13–16]. Below the threshold, the system relaxes to a single fixed point in concentration space, with no coherent oscillations at all. In some systems, the precision may be still well below the theoretical bound as the system approaches this threshold. In these cases, the precision will never come close to the bound, for any size driving force.

As we show in §3, computational models of real chemical oscillations typically fall into this regime, never approaching to within an order of magnitude of the bound. Macroscopic *in vitro* experiments on the KaiABC system perform even worse, remaining many orders of magnitude below the bound. Previous theoretical work suggests that the performance could be improved by increasing the number of reactions per cycle at fixed entropy production and by making the reaction rates more uniform [4,11,17]. In §2, we elaborate on these ideas, introducing an effective number of states per cycle and showing how the relationship of this quantity to the location of the phase transition threshold controls the minimum distance to the bound. In §4, we introduce a new model based on these design principles, with nearly uniform transition rates in the steady state and with a tunable number of reactions per cycle. We show that this model

approaches the optimal precision as the number of reaction steps per cycle grows.

2. Effective number of states and critical entropy production control distance to thermodynamic bound

As illustrated in figure 1a, a KaiC monomer has two phosphorylation sites, one on a threonine residue (T) and one on a serine (S), giving rise to four possible phosphorylation states [1]. The monomer also has two ATP-binding pockets, and forms hexamers that collectively transition between two conformational states [18,19]. All these features are important for the dynamics of the system and they have been incorporated into a thermodynamically consistent computational model that correctly reproduces the results of experiments performed with purified components [2]. In particular, the ATP hydrolysis rate in one of the binding pockets has been shown to be essential for determining the period of the circadian rhythms [18–20]. Although our simulations will use the detailed model just mentioned, the simplified schematic in figure 1 highlights only the features of the model that we will explicitly discuss: the fact that each molecule can execute a directed cycle among several internal states, and the fact that the state of one molecule affects transition rates of the others.

Fluctuations in the time required for a simplified model of a KaiC hexamer to traverse the reaction cycle have recently been studied in [6]. But the biological function of this clock demands more than precise oscillations of isolated molecules; rather, it has evolved to generate oscillations in the *concentrations* of various chemical species. The concentrations are global variables, which simultaneously affect processes throughout the entire cell volume. These global oscillations can still be described by a Markov process on a set of discrete states, but with a very different topology from the unicyclic network of an isolated monomer. For a well-mixed system, each state can be labelled by a list of copy numbers of all molecules in the reaction volume as shown in figure 1*b*, with each distinct internal state counted as a different kind of molecule.

In the KaiC system, molecules in one of the phosphorylation states can suppress further phosphorylation of other molecules, by sequestering the enzyme (KaiA) required to catalyse the phosphorylation. This mechanism can stably synchronize the progress of all the molecules around the phosphorylation cycle, slowing down the ones that happen to run too far ahead of the rest. This is crucial for the maintenance of sustained oscillations in the concentration of free KaiA and of each of the four forms of KaiC. Figure 1*c* shows a sample trajectory of the concentration of one of the KaiC phosphorylation states in the detailed computational model mentioned above [2].

Unlike the cycles of an idealized mechanical clock, the period τ_1 of these oscillations is subject to random fluctuations, due to the stochastic nature of the underlying chemical reactions. The precision can be quantified by considering an ensemble of identical reaction volumes that are initially synchronized. The histogram of times τ_n for each molecule to complete n cycles will widen as n increases and the clocks lose their initial synchronization, as illustrated in figure 1*d*. When the width exceeds the mean period $T \equiv \langle \tau_1 \rangle$, the clocks are totally desynchronized. This leads to a natural measure of the precision of the clock in terms of the number of coherent cycles \mathcal{N} that take place before the synchronization is destroyed.

To measure \mathcal{N} in a systematic way, we first note that the variance $\text{var}(\tau_n) = Dn$, for some constant of proportionality D , as illustrated in figure 1*e*. This is exactly true in a renewal process [21], such as the isolated KaiC monomer, where each period is an independent random variable (see [11]). It remains asymptotically valid for arbitrarily complex models in the large n limit, as long as the correlation time is finite. The number of cycles required for the width $\sqrt{\text{var}(\tau_n)}$ of the distribution to reach the average period T is therefore given by

$$\mathcal{N} \equiv \frac{T^2}{D}. \quad (2.1)$$

Any chemical oscillator must be powered by a detailed-balance-breaking thermodynamic driving force that generates a positive average rate of entropy production \dot{S} . The number of coherent cycles is subject to a universal upper bound as a function of \dot{S} , holding for arbitrarily complex architectures [4–6,8–10,22]. The bound says that \mathcal{N} is never greater than half the entropy production per cycle $\Delta S \equiv \dot{S}T$ (setting Boltzmann's constant $k_B = 1$ from here on) [7],

$$\mathcal{N} \leq \frac{\Delta S}{2}. \quad (2.2)$$

The validity of this bound depends on the proper definition of \mathcal{N} , which in our formulation also depends on the definition of τ_n . Determining τ_n is a subtle matter for systems of interacting molecules. Our solution is presented in detail in appendix A, but it always roughly corresponds to the peak-to-peak distance in figure 1*c*.

As $\Delta S \rightarrow \infty$, equation (2.2) says that \mathcal{N} is also allowed to become arbitrarily large. But as the entropy released in the reactions coupled to the driving force increases, detailed balance implies that the reverse reaction rates tend towards zero. Once the reverse rates are negligible compared with the other time scales of the problem, these reactions can be ignored, and further changes in ΔS produce no effect. In any given biochemical model, therefore, \mathcal{N} approaches some finite value as $\Delta S \rightarrow \infty$ (as was already noted in [3]), which depends on the network topology and the rest of the reaction rates. We can see this in our detailed computational model in figure 2. For unicyclic networks in particular, where the topology is a single closed cycle like the isolated KaiC monomer, the maximum possible value for this asymptote is the number of states N [12,23]. By analogy, we will refer to the $\Delta S \rightarrow \infty$ limit of \mathcal{N} for any model as the effective number of states per cycle $N_{\text{eff}} \equiv \lim_{\Delta S \rightarrow \infty} \mathcal{N}$. For an oscillator built from coupled cycles of internal states, such as the KaiC system, N_{eff} reaches its maximum value when the dynamics constrain the oscillations to a single path through concentration space, and when all reaction rates along this path are equal. In this case, the dynamics are equivalent to a single ring of NM states, where N is the number of internal states per molecule and M is the number of molecules. This upper bound on N_{eff} can be easily computed for any model or experiment from the basic knowledge of the component parts.

In all five models we will analyse below, \mathcal{N} monotonically increases as a function of ΔS . The existence of the finite $\Delta S \rightarrow \infty$ limit thus implies that \mathcal{N} can only approach the thermodynamic bound of equation (2.2) when $\Delta S < \Delta S_b \equiv 2N_{\text{eff}}$. But the collective oscillations of these models also exhibit a non-equilibrium phase transition as a function of ΔS , whose critical behaviour has recently been studied [3,24]. In the thermodynamic limit, the inverse precision $1/\mathcal{N}$ diverges as ΔS approaches a critical value ΔS_c from above, in a way that depends on the architecture of the reaction network. Below ΔS_c , there are no collective oscillations, and the concentrations relax to a single fixed point. Since the oscillations cease to exist below ΔS_c , the bound is only relevant for $\Delta S > \Delta S_c$. Combining these two observations, we see that models with $\Delta S_b < \Delta S_c$ can never approach the thermodynamic bound.

3. Models of real chemical oscillators severely underperform the bound

Cao *et al.* [3] recently measured \mathcal{N} as a function of ΔS in computational models of four representative chemical clock architectures: activator–inhibitor, repressilator, Brusselator and the glycolysis network. The data for all four models produced an acceptable fit to a four-parameter phenomenological equation, which is reproduced in appendix B along with the parameter values obtained by Cao *et al.* for each model. In figure 3, we plot these phenomenological curves and the thermodynamic bound of equation (2.2). We also obtained \mathcal{N} and ΔS for a detailed model of the KaiC system based on [2] as

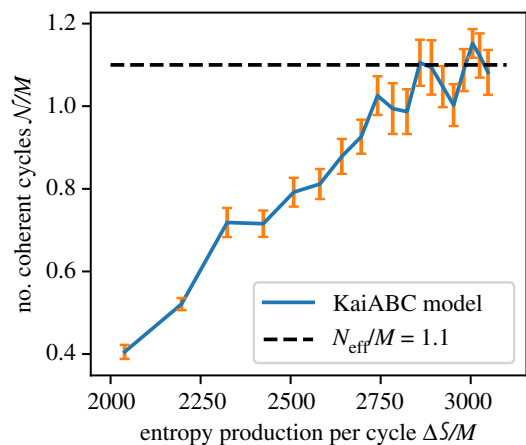


Figure 2. Number of coherent oscillations saturates as $\Delta S \rightarrow \infty$. The number of coherent cycles \mathcal{N} is plotted as a function of the entropy production per cycle ΔS for the KaiC model discussed in figure 1. Both axes are scaled by the system size M , so that all quantities are molecular-scale values. Error bars are ± 1 s.d., estimated with the bootstrap procedure described in appendix F. The black dotted line is the estimated $\Delta S \rightarrow \infty$ limit $\mathcal{N} = N_{\text{eff}}$. See appendix F and [2] for model parameters. (Online version in colour.)

described in appendix C, with the parameters obtained in that paper by extensive comparison with experimental data, for 20 values of the ATP/ADP ratio.

The values of N_{eff} , ΔS_b and ΔS_c can be estimated directly from figure 3, by noting where each curve saturates and where it drops to zero. Both axes are scaled by the system size M , which equals the number of KaiC hexamers for the KaiC model, and the number of kinases in the activator–inhibitor model. The other three models lack a direct physical interpretation of M , since there are no conserved molecular species, but it still defines a generic molecular scale. For any physically reasonable model, \mathcal{N} is expected to be an extensive parameter, proportional to M , as is ΔS . This has been confirmed numerically for a number of models, and appears to break down significantly only in the immediate vicinity of the critical point [3,15,24]. The models plotted here have $N_{\text{eff}}/M \approx 2$, which is reasonable for molecules that only have a few internal states and highly non-uniform reaction rates.

But $N_{\text{eff}}/M \approx 2$ implies that $\Delta S_b/M \approx 4$, which means that the entropy production per molecule per cycle must be less than 4 for the thermodynamic bound to become relevant. This is a very small number even by biochemical standards, equal to the entropy change from forming four hydrogen bonds between protein residues in solution. The activator–inhibitor, Brusselator and glycolysis models have phase transitions at $\Delta S_c/M$ values of 360, 100.4 and 80.5, respectively, under the parameter choices of [3]. They all exceed $\Delta S_b/M$ by at least an order of magnitude, guaranteeing that the precision can never come close to the thermodynamic bound. The KaiC model appears to have $\Delta S_c/M \sim 1000$ and $N_{\text{eff}}/M = 1.1$, so that ΔS_c exceeds ΔS_b by more than two orders of magnitude. The only model with $\Delta S_b > \Delta S_c$ is the repressilator model, where $\Delta S_c/M = 1.75$ and $\Delta S_b \approx 4$. But, even here, the critical fluctuations begin to severely degrade the precision when ΔS is still much greater than ΔS_b .

Estimates of \mathcal{N} can also be extracted directly from experiments, as shown by Cao *et al.* [3] for the *in vitro* reconstitution of the KaiC system with purified components

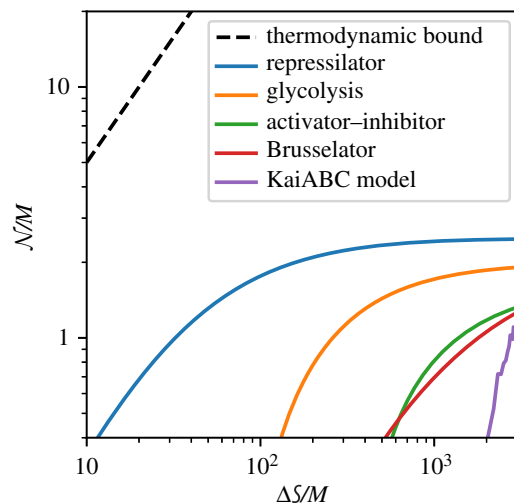


Figure 3. Models of collective oscillations compared with thermodynamic bound. Same as figure 2, but including all four models studied in [3]. The black dotted line is the thermodynamic bound $\mathcal{N} = \Delta S/2$. Curves for the first four models are phenomenological fits obtained in [3]. (Online version in colour.)

in a macroscopic reaction volume. They analysed time-series data from a set of experiments at different ATP/ADP ratios and fitted their phenomenological equation to describe \mathcal{N} as a function of this ratio. As we show in appendix C, this fit implies that $\Delta S_c/M \sim 1000$, consistent with our model results. The $\Delta S \rightarrow \infty$ asymptote of the fit, however, gives $N_{\text{eff}}/M \sim 10^{-11}$, which is astronomically small compared with the model prediction $N_{\text{eff}}/M \approx 1$. This surprising result reflects the fact that the dominant sources of uncertainty in these macroscopic experiments are fluctuations in temperature and other environmental perturbations, rather than the intrinsic stochasticity of the reaction kinetics. Since these fluctuations are independent of the system size, their effect is inflated when we divide by the number of hexamers $M \sim 10^{14}$ in a 100 μl reaction volume at 1 μM concentration. The only way to observe the effect of intrinsic stochasticity in such a noisy environment is to decrease the reaction volume. Assuming that the minimum contribution of the external noise to \mathcal{N} remains fixed at the value of 500 found in the experiments, and that the intrinsic contribution is of order $N_{\text{eff}} \approx M$ as given by the model, we find that $M = 500$ hexamers is the system size at which the intrinsic fluctuations become detectable. At 1 μM concentration of hexamers, the corresponding reaction volume is about 1 μm^3 , the size of a typical bacterial cell.

Because of this separation of scales, the apparent divergence of $1/\mathcal{N}$ at a critical value of the ATP/ADP ratio in the experiments is probably due to the expected divergence of susceptibility at the critical point, which makes the oscillation period increasingly sensitive to environmental fluctuations as the ATP/ADP ratio is reduced. In any case, this analysis suggests that an important design consideration for oscillators in living systems is robustness against external perturbations, as recently explored in [25–27].

4. Toy model with variable number of states can saturate the bound

The failure of all five of these models to approach the thermodynamic bound raises the question of whether it is possible, in

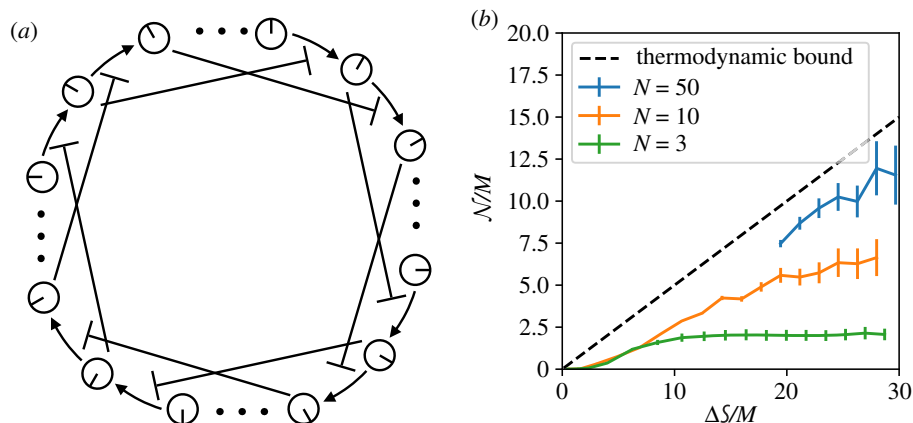


Figure 4. Symmetric toy model compared with thermodynamic bound. (a) Schematic of toy model inspired by the KaiC system. Each protein has N distinct conformations whose transitions are arranged in a ring topology with a net clockwise drift. Each protein suppresses the transition rate for proteins further along in the circulation around the ring. (b) N/M versus $\Delta S/M$ in this model for three different values of the number of internal states N . Error bars are ± 1 s.d., estimated with the same bootstrap procedure used in figure 2. See appendix E for model details and parameters. (Online version in version)

principle, for any chemical oscillator to simultaneously achieve a large enough N_{eff}/M and small enough $\Delta S_c/M$ to do so. In a simple unicyclic network of N states, $N_{\text{eff}} = N$ when all the reaction rates in the forward direction are equal, and so one can always approach arbitrarily close to the bound by increasing N . But a chemical oscillator cannot have uniform rates, since the transition rates of each molecule have to change based on the states of the others in order to achieve collective oscillations. Furthermore, it is not known how changing the number of internal states affects $\Delta S_c/M$, and so it is not obvious whether $\Delta S_c < \Delta S_b$ is achievable at all.

To answer this question, we devised a new model loosely inspired by the KaiC model of figure 1, with M interacting molecules each containing N distinct internal states. Molecules in any one of these states suppress the transition rates for other molecules that are further ahead in the cycle, as illustrated in figure 4 and described in detail in appendix E. All internal states have the same energy, and each reaction carries the same fraction of the total thermodynamic force.

In this highly symmetrized model, $\Delta S_c/M$ can easily be reduced to between 2 and 3 by choosing a sufficiently high coupling strength, as shown in appendix E. At the same time, N_{eff}/M scales linearly with N and it can be made arbitrarily large by increasing this parameter. In figure 4, we plot N/M versus $\Delta S/M$ for three different values of N and show that the data do indeed approach the thermodynamic bound as N increases. This extends the validity of design principles obtained for unicyclic networks in various contexts to these collective dynamics: the rates should be made as uniform as possible, while the number of internal states is made as large as possible at fixed thermodynamic driving force [6,11,28].

5. Discussion

The thermodynamic uncertainty relation is a powerful result with impressive universality. It has been widely assumed that the relation should have some relevance for the evolution of biochemical oscillators. Based on data from experiments and extensive simulations with realistic parameters, we have argued that these oscillators typically underperform the bound by at least an order of magnitude. From a

thermodynamic perspective, they are free to evolve higher precision without increasing their dissipation rate.

We have also derived a simple criterion for estimating how closely a given oscillator can approach the thermodynamic bound, in terms of an effective number of states N_{eff} and the entropy production per cycle ΔS_c at the onset of oscillatory behaviour. For an oscillator composed of M identical molecules with N internal states, globally coupled through mass-action kinetics, we noted that $N_{\text{eff}} \leq NM$, with equality only when all the cycles are perfectly synchronized, and when all reactions that actually occur have identical rates. Assuming that the number of coherent periods \mathcal{N} is monotonic in the entropy production per cycle ΔS , the thermodynamic bound can only be approached when $NM \geq N_{\text{eff}} \gg \Delta S_c$.

To show that this criterion can, in principle, be satisfied by emergent oscillations of molecular concentrations, we devised a toy model that oscillates with less than $3 k_B T$ of free energy per molecule per cycle and can contain an arbitrary number of internal states per molecule. But it is hard to imagine a biochemically plausible mechanism for sustained oscillations powered by the free energy equivalent of three hydrogen bonds per molecule per cycle. Certainly this could not be a phosphorylation cycle, since acquiring a phosphate group from ATP and releasing it into the cytosol dissipates about $20 k_B T$ under physiological conditions [29]. Furthermore, we noted that all the five models we analysed have an effective number of internal states per molecule N_{eff}/M of around 2. This number may be constrained by a trade-off with the complexity of the oscillator. For KaiC, the inter-molecular coupling required for collective oscillations is mediated by sequestration of KaiA at a particular point in the cycle. Implementing the symmetric N -state cycle of figure 4a in this way would require each transition to be catalysed by a different molecule, and for each of these molecules to be selectively sequestered by the state a quarter-cycle behind the transition. Finally, biochemical oscillators are subject to other performance demands that may be more important than the number of coherent cycles of the free-running system. In particular, an oscillator's ability to match its phase to an external signal (e.g. the day/night cycle of illumination intensity) is often essential to its

biological function, placing an independent set of constraints on the system's architecture [27,30]. Entrainment efficiency has recently been shown to increase with dissipation rate in some models even when the free-running precision has saturated, providing an impetus for increasing the entropy production per cycle beyond what is required to achieve $\mathcal{N} = \mathcal{N}_{\text{eff}}$ [31].

Data accessibility. All simulation data and scripts for regenerating figures can be found at <https://github.com/robertvsiii/kaic-analysis>.

Authors' contributions. R.M. and J.M.H. conceived of the study. R.M. performed analytical calculations and simulations, and drafted the manuscript. W.C. and J.M.H. performed simulations and critically revised the manuscript.

Competing interests. We declare we have no competing interests.

Funding. R.M. acknowledges Government support through NIH NIGMS grant no. 1R35GM119461. J.M.H. is supported by the Gordon and Betty Moore Foundation as a Physics of Living Systems Fellow through grant no. GBMF4513.

Acknowledgements. The authors thank J. Paijmans for his help with adjustments to the KaiC simulation, and Y. Cao for the valuable discussions on the technical details of reference [3]. The computational work reported on in this paper was performed on the Shared Computing Cluster, which is administered by Boston University's Research Computing Services.

Appendix A. Measuring the stochastic period

The definition of the number of coherent cycles \mathcal{N} depends on a prior notion of the n -cycle completion time τ_n . In a unicyclic transition network, this time can be straightforwardly defined in terms of the integrated current J through an arbitrarily chosen transition in the network. Each time the transition is executed in the forward direction, J increases by 1, and each time it is executed in the reverse direction, J decreases by 1. The n -cycle completion time τ_n is then naturally defined as the time when the system first reaches $J = n$, given that it was initialized in the state immediately adjacent to the measured transition in the positive direction [7,11,21].

For a chemical oscillator, the definition is not so clear. One common approach is to fit the autocorrelation function of some observable to a sine wave with exponentially decaying amplitude. If the autocorrelation function exactly fits this functional form, then \mathcal{N} can be obtained from the ratio of the decay time to oscillation period via a numerical conversion factor [3]. One can also evaluate the ratio of imaginary to real parts of the leading eigenvalue of the rate matrix for the master equation of the dynamics, which gives the same result when all the other eigenvalues are much smaller in amplitude [6,13,16]. While this ratio is conjectured to be bounded by the thermodynamic driving force powering the oscillations, it is not the approach we study here [6]. Instead, we note that the value of \mathcal{N} generated by these preceding procedures only satisfies the hypotheses of the thermodynamic uncertainty relation under the specific conditions of a perfectly sinusoidal autocorrelation function (see [6]). For our analysis of the KaiC model and our new toy model, we instead use a definition of τ_n that treats the oscillations in the full concentration space as one large cycle.

Figure 5 illustrates our procedure. We started by projecting the state of the system from the high-dimensional concentration space to two dimensions. We projected onto the plane that captured the largest percentage of the total variation in system state over a cycle, using a principal

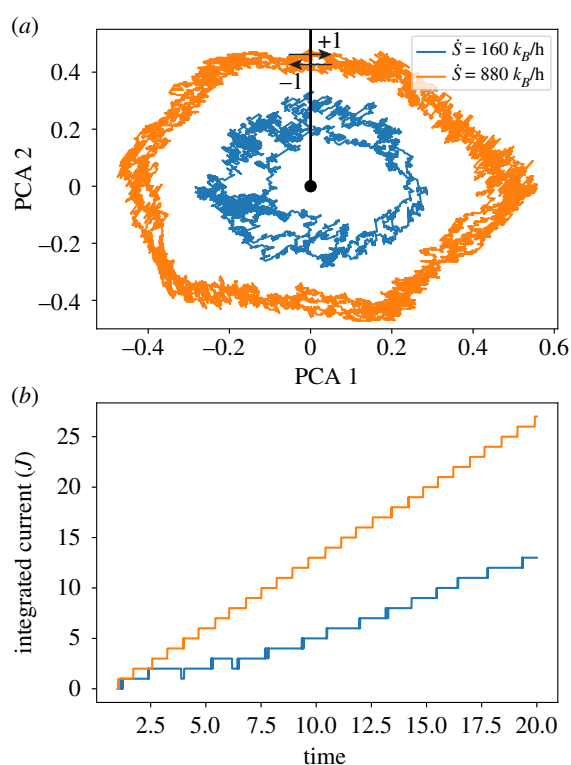


Figure 5. Measuring the n -cycle completion time τ_n . (a) Trajectories of the toy model of figure 4 with $N = 6$ for two different values of the thermodynamic driving force, projected onto their first two principal components. A cut from the origin along the positive vertical axis provides the criterion for cycle completion. (b) Integrated current (net number of completed cycles) J as a function of time for the same two trajectories. The first-passage time for a net increase n in the integrated current defines the n -cycle completion time τ_n . (Online version in colour.)

component analysis (PCA) of a trajectory containing at least one full cycle (using the Python package scikit-learn [32]). In this plane, the oscillating trajectories describe a noisy ring, as shown in figure 5a. Because the ring has a finite width, we cannot select a single transition to count the integrated current J . Instead, we draw a half-line starting from the middle of the ring, representing a half-hyperplane in the full state space, and include all the transitions cut by this hyperplane in the current. Each time the line is crossed in the clockwise direction, J increases by 1, and each time it is crossed in the counterclockwise direction, J decreases by 1. Sample traces of $J(t)$ are plotted in figure 5b. The n -cycle completion time τ_n can now be defined as before, measuring the first-passage time for reaching $J = n$. These definitions fulfil the hypotheses of the thermodynamic uncertainty relations for currents and for first-passage times [4,7,8]. In the notation of [7,8], they correspond to setting $d(y, z) = 1$ for all transitions from y to z cut by the hyperplane, and $d(y, z) = 0$ for all other transitions. Note that the uncertainty relations are obtained in the $J \rightarrow \infty$ limit, where initial conditions are irrelevant, but for our numerical analysis we chose special initial conditions that gave rapid convergence to the asymptotic form. Specifically, we employed a conditional steady-state distribution over states adjacent to the hyperplane. This was achieved by running the simulation longer than the relaxation time, and then starting the counter at $J = 0$ the next time the hyperplane was crossed.

Table 1. Parameters of phenomenological fits from [3].

model	A	B	$\Delta S_c/M$	α
activator–inhibitor	0.6	380	360	−0.99
repressilator	0.4	25.9	1.75	−1.1
Brusselator	0.5	846	100.4	−1.0
glycolysis	0.5	151.4	80.5	−1.1

Note that A controls the $\Delta S \rightarrow \infty$ asymptote and is equal to N_{eff}^{-1} .

Appendix B. Phenomenological fits from [3]

The extensive numerical simulations performed by Cao *et al.* [3] on four different models of chemical oscillators can be summarized by the parameters of a phenomenological fitting function

$$\frac{\mathcal{N}}{M} = \left[A + B \left(\frac{\Delta S - \Delta S_c}{M} \right)^\alpha \right]^{-1}, \quad (\text{B } 1)$$

with four parameters A , B , ΔS_c and α . The exponent α is always negative, so \mathcal{N}/M goes to zero as $\Delta S \rightarrow \Delta S_c$. (To convert from their notation to ours, use $V \rightarrow M$, $W_0 \rightarrow B$, $C \rightarrow A$, $W_c \rightarrow \Delta S_c/M$, $D/T \rightarrow \mathcal{N}^{-1}$.) The parameters for these fits are given in the figure captions of [3] and are reproduced in table 1.

Appendix C. Analysis of experimental data

Cao *et al.* [3] also analyse experimental data on the KaiC system, extracting the ratio of decay time to oscillation period from fits to the autocorrelation function at different values of the ATP/ADP ratio. They fit equation (B 1) to the data, but using $\ln([\text{ATP}]/[\text{ADP}])$ instead of $\Delta S/M$, and without converting the autocorrelation ratio to \mathcal{N} or dividing by volume.

To compare these results with the thermodynamic bound, we first had to convert from the logarithm of the ATP/ADP ratio to entropy production per cycle. In the text, they estimate that the critical value $\ln([\text{ATP}]/[\text{ADP}])_c = -1.4$ obtained from the fit corresponds to an entropy production per ATP hydrolysis event of 10.6. Combining this with the measurement from [20] of 16 hydrolysis events per cycle per KaiC monomer, we find a critical entropy production per cycle per hexamer of $\Delta S_c/M \approx 10.6 \times 16 \times 6 \approx 1020$.

Next, we had to convert the observed ratio of decay time/period to \mathcal{N}/M . Since the autocorrelation function was well fitted by an exponentially decaying sinusoid, we applied the corresponding conversion factor $\mathcal{N} = 2\pi^2(\tau/T)$, where T is the period and τ is the decay time [3, eqn (2)]. We then estimated the number of hexamers $M \approx 3 \times 10^{13}$ using the KaiC monomer concentration of 3.4 μM reported with the original publication of the data [33,34], and the typical reaction volume in a 96-well plate of 100 μL . With these two conversions, we found that the $\Delta S \rightarrow \infty$ value of \mathcal{N}/M was 2×10^{-11} .

Appendix D. Thermodynamically consistent KaiC model

Paijmans *et al.* [2] have recently developed a mechanistically explicit computational model of the KaiC oscillator. This model is particularly interesting from the theoretical point

of view because it captures the extremely large dimensionality characteristic of real biochemical systems. Each KaiC hexamer contains six KaiC proteins, each of which contains two nucleotide-binding sites and two possible phosphorylation sites ('S' and 'T' from figure 1). Each of these sites can be in one of two possible states (ATP-bound/ADP-bound or phosphorylated/unphosphorylated). Furthermore, the whole hexamer can be in an 'active' or an 'inactive' conformation. Thus each hexamer has $(2 \times 2 \times 2 \times 2)^6 \times 2 = 2^{25}$ possible internal states. As noted in the main text, the state space for a well-mixed chemical system is the vector of concentrations of all molecular types. For the Paijmans *et al.* KaiC model, this vector therefore lives in a space of dimension $2^{25} = 3.4 \times 10^7$.

The original implementation of this model in [2] lacked the reverse hydrolysis reaction $\text{ADP} + \text{P} \rightarrow \text{ATP}$, which never spontaneously happens in practice under physiological conditions. To obtain full thermodynamic consistency, we added this reaction to the model with the assistance of one of the original authors. This required introducing a new parameter ΔG_0 , the free energy change of the hydrolysis reaction at standard concentrations. For all the simulations analysed here, we chose ΔG_0 and the concentration of inorganic phosphate [Pi] such that

$$\frac{[\text{Pi}]_0}{[\text{Pi}]} e^{-\Delta G_0} = 10^8. \quad (\text{D } 1)$$

In other words, the entropy generated during a single hydrolysis reaction when nucleotide concentrations are equal ($[\text{ATP}] = [\text{ADP}]$) is $\Delta S_{\text{hyd}} = \ln 10^8 \approx 18.4$.

Since the steady-state supply of free energy in this model comes entirely from the fixed non-equilibrium concentrations of ATP and ADP, we can measure the average rate of entropy production \dot{S} by simply counting how many ATP molecules are hydrolysed over the course of a long simulation, multiplying by ΔS_{hyd} , and dividing by the total time elapsed in the simulation.

All parameters other than ΔG_0 are described and tabulated in the original publication [2], and the only parameter altered during our simulations was the ATP/ADP ratio.

The revised C code and scripts for generating data can be found on GitHub at <https://github.com/robertvsiii/kaic-analysis>.

Appendix E. Symmetric toy model

We also developed our own abstract toy model to isolate the operating principles of the KaiC oscillator and to check whether the thermodynamic bound could be saturated by a collective oscillator with the right design.

Consider a molecule with N states, as sketched in figure 6a. Transitions are allowed from each state to two other states, such that the network of transitions has the topology of a ring. The rates for 'clockwise' and 'counterclockwise' transitions around this ring are $k^+ = Nk$ and $k^- = Nk e^{-A/N}$, respectively, where A is the cycle affinity. Under these definitions, the total entropy produced when one ring executes a full cycle is always equal to A .

Now consider M copies of this molecule in the same well-mixed solution with M_i copies at position i , where i increases in the 'clockwise' direction from 1 to N . We can couple their dynamics together by allowing the bare rate k to vary around the ring. Specifically, we make the bare

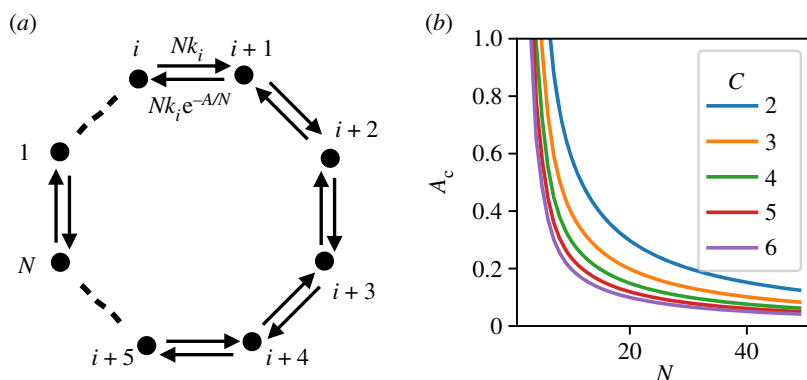


Figure 6. Toy model with variable number of internal states. (a) Transition rates for a single molecule with N internal states. Multiple copies of the molecule are coupled together kinetically, by making k_i depend on the fraction of molecules f_j in each state. (b) Dependence of critical affinity A_c on N for different values of the coupling C . (Online version in colour.)

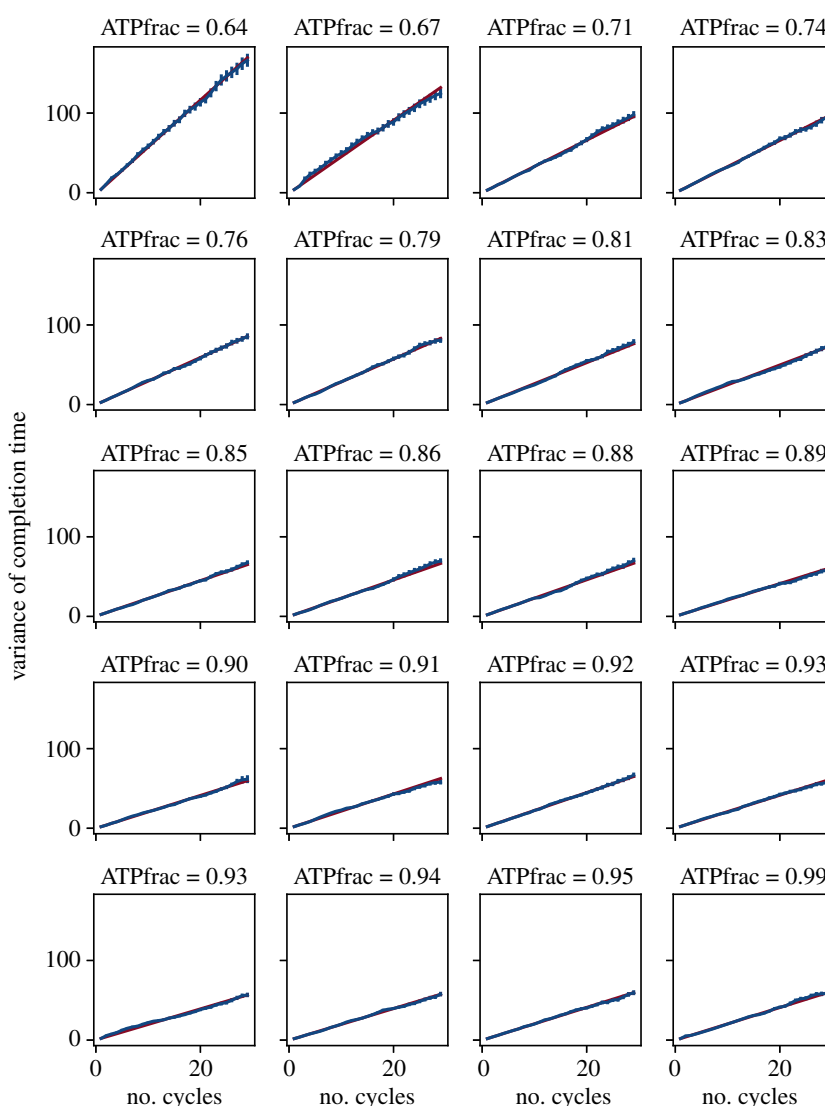


Figure 7. Estimating \mathcal{N} from KaiC simulations. Each panel shows the estimated variance and bootstrapped 64% confidence intervals in the n -cycle first-passage time τ_n . Straight black lines are linear fits, whose slopes provide the values of D used in the computation of \mathcal{N} for the main text figures. (Online version in colour.)

rate k_i for transitions between states i and $i + 1$ depend on the occupancy fractions $f_j = M_j/M$ of all N states,

$$k_i = \exp\left(-C \sum_j f_j \sin\left[\frac{2\pi(i-j)}{N}\right]\right). \quad (\text{E } 1)$$

The constant C controls the strength of the coupling, and the rate for a uniform distribution over states is 1. This

dependence of the rates on the f_j mimics the effect of KaiA sequestration in the KaiC system. Recall that high occupancy of the inactive conformation of KaiC causes KaiA to be sequestered, slowing down nucleotide exchange in other hexamers, as illustrated in figure 1. In this toy model, high occupancy of any one state slows down transition rates ahead of that state in the cycle, by up to a factor of e^{-C} for transitions a quarter-cycle ahead. Owing to the symmetry

of our model, high occupancy of a given state also speeds up transition rates behind that state.

The data for figure 4 were obtained with $C = 5$, for 18 values of A from 1 to 30. Note that $\Delta S/M \approx A$, since all M molecules execute approximately one cycle during a given oscillation period.

We simulated this model using a Gillespie algorithm with the reaction rates specified above. The Python code can be found in the GitHub repository <https://github.com/robertvsiii/kaic-analysis>.

In the limit of infinite system size, the dynamics become deterministic, and are described by the following set of N ODEs:

$$\frac{df_i}{dt} = f_{i-1}k_i^+ + f_{i+1}k_i^- - f_i(k_i^+ + k_{i-1}^-), \quad (\text{E } 2)$$

with $k_i^+ = Nk_i$ and $k_i^- = Nk_i e^{-A/N}$. These equations always have a fixed point at the uniform state where $f_i = (1/N)$ for all i . The linearized dynamics around the uniform state can be written as follows:

$$\begin{aligned} \frac{d\delta f_i}{dt} = & \frac{1}{N} \sum_j \left[\frac{\partial k_{i-1}^+}{\partial f_j} + \frac{\partial k_i^-}{\partial f_j} - \left(\frac{\partial k_i^+}{\partial f_j} + \frac{\partial k_{i-1}^-}{\partial f_j} \right) \right] \delta f_j \\ & + (\delta f_{i-1} - \delta f_i)N + (\delta f_{i+1} - \delta f_i)N e^{-A/N} \\ = & \sum_j K_{ij} \delta f_j, \end{aligned} \quad (\text{E } 3)$$

where the last line defines the matrix K_{ij} . Oscillating solutions are possible when K_{ij} acquires an eigenvalue with a positive real part, making this fixed point unstable. In figure 6*b*, we plot the critical affinity A_c , where these positive real parts first appear, as a function of the number of internal states N . We confirmed that the dominant pair of eigenvalues contains nonzero imaginary parts at $A = A_c$ for all points plotted, so that the transition is a true Hopf bifurcation to an oscillatory phase.

References

- Nakajima M, Imai K, Ito H, Nishiwaki T. 2005 Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* **308**, 414–415. (doi:10.1126/science.1108451)
- Paijmans J, Lubensky DK, ten Wolde PR. 2017 A thermodynamically consistent model of the post-translational Kai circadian clock. *PLoS Comput. Biol.* **13**, e1005415. (doi:10.1371/journal.pcbi.1005415)
- Cao Y, Wang H, Ouyang Q, Tu Y. 2015 The free-energy cost of accurate biochemical oscillations. *Nat. Phys.* **11**, 772–778. (doi:10.1038/nphys3412)
- Barato AC, Seifert U. 2015 Thermodynamic uncertainty relation for biomolecular processes. *Phys. Rev. Lett.* **114**, 158101. (doi:10.1103/PhysRevLett.114.158101)
- Barato AC, Seifert U. 2016 Cost and precision of Brownian clocks. *Phys. Rev. X* **6**, 041053. (doi:10.1103/PhysRevX.6.041053)
- Barato AC, Seifert U. 2017 Coherence of biochemical oscillations is bounded by driving force and network topology. *Phys. Rev. E* **95**, 062409. (doi:10.1103/PhysRevE.95.062409)
- Gingrich TR, Horowitz JM. 2017 Fundamental bounds on first passage time fluctuations for currents. *Phys. Rev. Lett.* **119**, 170601. (doi:10.1103/PhysRevLett.119.170601)
- Gingrich TR, Horowitz JM, Perunov N, England JL. 2016 Dissipation bounds all steady-state current fluctuations. *Phys. Rev. Lett.* **116**, 120601. (doi:10.1103/PhysRevLett.116.120601)
- Horowitz JM, Gingrich TR. 2017 Proof of the finite-time thermodynamic uncertainty relation for steady-state currents. *Phys. Rev. E* **96**, 020103. (doi:10.1103/PhysRevE.96.020103)
- Pietzonka P, Barato AC, Seifert U. 2016 Universal bounds on current fluctuations. *Phys. Rev. E* **93**, 052145. (doi:10.1103/PhysRevE.93.052145)
- Wierenga H, ten Wolde PR, Becker NB. 2018 Quantifying fluctuations in reversible enzymatic cycles and clocks. *Phys. Rev. E* **97**, 042404. (doi:10.1103/PhysRevE.97.042404)
- Marsland R, England J. 2018 Limits of predictions in thermodynamic systems: a review. *Rep. Prog. Phys.* **81**, 016601. (doi:10.1088/1361-6633/aa9101)
- Nguyen B, Seifert U, Barato AC. 2018 Phase transition in thermodynamically consistent biochemical oscillators. *J. Chem. Phys.* **149**, 1. (doi:10.1063/1.5032104)
- Herpich T, Thingna J, Esposito M. 2018 Collective power: minimal model for thermodynamics of nonequilibrium phase transitions. *Phys. Rev. X* **8**, 031056. (doi:10.1103/physrevx.8.031056)
- Lee S, Hyeon C, Junghyo J. 2018 Thermodynamic uncertainty relation of interacting oscillators in synchrony. *Phys. Rev. E* **98**, 032119. (doi:10.1103/PhysRevE.98.032119)
- Qian H, Qian M. 2000 Pumped biochemical reactions, nonequilibrium circulation, and stochastic resonance. *Phys. Rev. Lett.* **84**, 2271. (doi:10.1103/PhysRevLett.84.2271)
- Pietzonka P, Kleinbeck K, Seifert U. 2016 Extreme fluctuations of active Brownian motion. *New J. Phys.* **18**, 052001. (doi:10.1088/1367-2630/18/5/052001)
- Phong C, Markson JS, Wilhoite CM, Rust MJ. 2013 Robust and tunable circadian rhythms from

Note that, in the limit $A \rightarrow \infty$, $N \rightarrow \infty$, this model becomes identical to the irreversible limit of a fully connected driven XY model.

Appendix F. Simulations and analysis

To measure \mathcal{N} and ΔS in the KaiC model and our new toy model, we generated an ensemble of trajectories for each set of parameters. Each ensemble of the KaiC model contained 1200 trajectories, while each toy model ensemble contained 1120 trajectories. Before collecting data, we initialized each trajectory by running the dynamics for longer than the empirically determined relaxation time of the system, in order to obtain a steady-state ensemble.

After projecting each trajectory onto the first two principal components and computing the n -cycle first-passage times as described above, we obtained the variance in τ_n as a function of n for each ensemble. We computed bootstrapped 64% confidence intervals for the estimate of the variance using the Python module 'bootstrapped', available at <https://github.com/facebookincubator/bootstrapped>. These data are plotted for all the KaiC ensembles in figure 7. The data are well fitted by a straight line even for low n in each of the plots. We obtained the slope D of these lines using a weighted least-squares fit, also shown in the figure, with the weights provided by the inverse of the bootstrapped confidence intervals.

We used these confidence intervals to obtain the bootstrap estimate for the uncertainty in D . The size of the confidence interval was proportional to n , as expected from a simple multiplicative noise model where the slope D is a random variable. The constant of proportionality yields an estimate for the standard deviation of the distribution from which D is sampled. We obtained this value for each data point with another least-squares linear fit and used it to set the size of the error bars in figures 2 and 4.

- differentially sensitive catalytic domains. *Proc. Natl Acad. Sci. USA* **110**, 1124–1129. (doi:10.1073/pnas.1212113110)
19. Tseng R *et al.* 2017 Structural basis of the day-night transition in a bacterial circadian clock. *Science* **355**, 1174–1180. (doi:10.1126/science.aag2516)
 20. Terauchi K, Kitayama Y, Nishiwaki T, Miwa K, Murayama Y, Oyama T, Kondo T. 2007 ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc. Natl Acad. Sci. USA* **104**, 16377. (doi:10.1073/pnas.0706292104)
 21. Ptaszyński K. 2019 First-passage times in renewal and nonrenewal systems. *Phys. Rev. E* **97**, 012127. (doi:10.1103/PhysRevE.97.012127)
 22. Dechant A, Sasa S-I. 2018 Current fluctuations and transport efficiency for general Langevin systems. *J. Stat. Mech: Theory Exp.* **2018**, 063209. (doi:10.1088/1742-5468/aac91a)
 23. David A, Larry S. 1987 The least variable phase type distribution is erlang. *Stoch. Models* **3**, 467–473. (doi:10.1080/153263487088807067)
 24. Nguyen M, Vaikuntanathan S. 2018 Dissipation induced transitions in elastic strings. (<http://arxiv.org/abs/1803.04368>)
 25. del Junco C, Vaikuntanathan S. 2018 High chemical affinity increases the robustness of biochemical oscillations. (<http://arxiv.org/abs/1808.04914>)
 26. Monti M, Lubensky DK, ten Wolde PR. 2018 Robustness of clocks to input noise. *Phys. Rev. Lett.* **121**, 078101. (doi:10.1103/PhysRevLett.121.078101)
 27. Pittayakanchit W, Lu Z, Chew J, Rust MJ, Murugan A. 2018 Biophysical clocks face a trade-off between internal and external noise resistance. *eLife* **7**, e37624. (doi:10.7554/eLife.37624)
 28. Lang AH, Fisher CK, Mora T, Mehta P. 2014 Thermodynamics of statistical inference by cells. *Phys. Rev. Lett.* **113**, 148103. (doi:10.1103/PhysRevLett.113.148103)
 29. Moran U, Phillips R, Milo R. 2010 Snapshot: key numbers in biology. *Cell* **141**, 1262–1262. (doi:10.1016/j.cell.2010.06.019)
 30. Chew J, Leypunskiy E, Lin J, Murugan A, Rust MJ. 2018 High protein copy number is required to suppress stochasticity in the cyanobacterial circadian clock. *Nat. Commun.* **9**, 3004. (doi:10.1038/s41467-018-05109-4)
 31. Fei C, Cao Y, Ouyang Q, Tu Y. 2018 Design principles for enhancing phase sensitivity and suppressing phase fluctuations simultaneously in biochemical oscillatory systems. *Nat. Commun.* **9**, 1434. (doi:10.1038/s41467-018-03826-4)
 32. Pedregosa F *et al.* 2011 Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.
 33. Phong C, Markson JS, Wilhoite CM, Rust MJ. 2013 Robust and tunable circadian rhythms from differentially sensitive catalytic domains. *Proc. Natl Acad. Sci. USA* **110**, 1124. (doi:10.1073/pnas.1212113110)
 34. Rust MJ, Golden SS, O'Shea EK. 2011 Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331**, 220–223. (doi:10.1126/science.1197243)