Mathematical Models in Biology: from Information Theory to Thermodynamics (Online)

Peter J. Thomas (Case Western Reserve University),
Michael Hinczewski (Case Western Reserve University),
Andrew Eckford (York University)

July 27–July 29, 2020

1 Overview of the Field

All living things, from the simplest bacteria to human beings, are made of cells. Fundamental understanding of living systems, both in health and in disease, depends on understanding the complex interactions among and within living cells. Multiple scientific disciplines have separately shed light on the problems of communication and organization in living systems. Biochemistry, bioinformatics and systems biology describe the basic ingredients of cells: DNA, RNA, proteins, lipids, and their interactions. Information theory, founded by Claude Shannon, provides a framework for quantifying the flow of information through any communications system, whether living or engineered (or both, as in the rapidly growing field of synthetic biology). Statistical thermodynamics, the branch of physics concerned with transformations among different forms of energy as well as with the physics of information, sets fundamental limits on the energetic price cells must pay for the information they sense (from each other, from the environment, and from their own DNA).

In the last five years, significant advances in statistical thermodynamics and the information theory of biological systems have set the stage for a deeper understanding of how cells process and organize information, make decisions, predict the future, and learn from the past. An essential link between these traditionally disparate fields is the language of mathematics, which provides a common framework within which researchers can understand each other across disciplines. The workshop on Mathematical Models in Biology: from Information Theory to Thermodynamics was planned to bring together leading experts and aspiring junior researchers from systems biology, statistical physics, information theory, and applied mathematics to develop the fundamental, linking ideas, to compare recent advances in their fields, and to establish new collaborations.

Stochastic Thermodynamics. The principles of classical thermodynamics have been established since the 19th century, including foundational notions such as the conservation of energy in its many forms (kinetic energy; gravitational and electrical potential energy; enthalpy U of chemical reactions; work), absolute temperature T, Boltzmann's entropy S, and Gibb's free energy $G = U - TS$. Boltzmann famously had his entropy formula $S = k \log W$ engraved as his epitaph, where k is Boltzmann's constant and W denotes *Wahrscheinlichkeit*, or probability. Decades later, Claude Shannon put the theory of communication systems on a firm mathematical basis by establishing the entropy $H = -\sum_i p_i \log p_i$ as the quantitative measure of *information* of a source producing the ith symbol with probability p_i. The formal similarity between the physical entropy S and Shannon's information measure H has spurred volumes of research aiming to elucidate their shared significance (if any). Particularly within theoretical biology, at the cellular or subcellular level,

one confronts processes that are naturally described one the one hand by the laws of chemistry, physics, and thermodynamics, and at the same time appear to function as teleological systems performing information-processing functions, e.g. communication, sensing, learning, or decision-making. The goal of the workshop was to advance interdisciplinary communication in this area.

2 Recent Developments and Open Problems

The last fifteen years have seen rapid advancement in the area of *stochastic thermodynamics*, led by the work of U. Seifert, cf. [Seifert (2008), Seifert (2019)], which provides the most promising intellectual framework yet for the analysis of information processing at cellular and subcellular scales within biology. Stochastic thermodynamics provides two conceptual advantages that are particularly relevant to biology: (i) It allows for a self-consistent thermodynamic description of arbitrarily small systems coupled to a thermal environment, for example individual biomolecules in solution. This is in contrast to conventional thermodynamics, which typically assumes both system and environment are macroscopic. (ii) It assigns thermodynamic quantities like entropy, work, and energy to individual time trajectories of a system. These trajectory-based definitions agree with the traditional ensemble-based definitions in the macroscopic limit, but they also reveal new physics: a variety of so-called "fluctuation theorems" that have been discovered over the last two decades. These theorems, which have been experimentally validated in biophysical systems, effectively generalize the second law of thermodynamics. They represent the most significant addition to our understanding of classical thermodynamics since the work of Boltzmann, Gibbs, and Maxwell in the 19th century. However the full implications of this novel physics for biological function, particularly information processing, are still being explored. The workshop provided a broad overview of the topic (through the opening talk of Udo Seifert), and in the subsequent talks illustrated the diverse applications of these ideas in biological systems.

3 Presentation Highlights

The workshop schedule, abbreviated to accommodate the online format, comprised five scientific talks (followed by ample time for discussion) and a virtual poster session.

3.1 Talk 1. Udo Seifert: From Stochastic Thermodynamics to Thermodynamic Inference

Udo Seifert (Univ. Stuttgart) set the stage by introducing stochastic thermodynamics, which apply to systems in which non-equilibrium is caused by mechanical or chemical forces, ambient solution provides a thermal bath of well-defined temperature T and chemical potential μ_i, and fluctuations are relevant due to small numbers of involved molecules [Collin et al (2005)]. He asked whether the same principles of thermodynamics that apply to the heat engines of the 19th century apply to molecular motors such as the F_1-ATPase rotor. The main idea, Seifert urged, is to take energy conservation and entropy production seriously along the individual trajectories [Seifert (2008)].

The setting for stochastic thermodynamics, reviewed in [Seifert (2012)], begins with a closed equilibrium system in contact with a thermal reservoir at fixed temperature $T \equiv \beta^{-1}$, nominally described by an ensemble of microstates ξ with energy $H(\xi)$. At equilibrium, state ξ occurs with probability $p_{\text{eq}}(\xi) = e^{-\beta H(\xi)}/Z$, normalized by the partition function $Z = \sum_\xi e^{-\beta H(\xi)}$. The (mean) internal energy $U = \sum_\xi p_{\text{eq}}(\xi) H(\xi)$, the entropy $S = \sum_\xi p_{\text{eq}}(\xi) \ln(1/p_{\text{eq}})$, and the free energy $F = U - TS$ are given by the classic thermodynamic relations $F = -\beta^{-1} \ln Z$, $U = \partial_\beta(\beta F)$, and $S = \beta^2 \partial_\beta F$. Thus $p_{\text{eq}}(\xi) = \exp(-\beta(H(\xi) - F))$.

One typically does not observe the system in a way that resolves individual microstates, but rather makes coarse-grained observations of an observable giving an ensemble of mesoscopic states $\{I\}$, with each microstate belonging to exactly one mesostate, $\xi \in I$. One can define the free energy $F(I)$, internal energy $U(I)$ and entropy $S(I)$ for mesoscopic states in terms of the conditional probabilities $p_{\text{eq}}(\xi \,|\, I)$. By observing long trajectories $I(t;\beta)$ at slightly different inverse temperatures β, one can estimate thermodynamic quantities associated with particular mesoscopic observables from experimental data. If the equilibration of the microstates within each mesostate is fast compared to the transition times between mesostates, then one

can recover an effective thermodynamics (first and second law) at the mesoscopic level. Along a stochastic mesoscopic trajectory $I(t)$ one has (stochastically) fluctuating quantities e.g. $U(I(t))$. Starting from an out-of-equilibrium initial probability distribution, $p_0(I) \neq p_{\rm eq}(I)$, one can compute the total entropy production, the heat dissipation, and the stochastic entropy production by comparing forward and (fictitious) backward-in-time trajectories. In order to consider driven systems, such as chemically driven molecular motors performing work against an applied load, one embeds the system of interest in a closed "supersystem" which is then partitioned into a core system and a connected reservoir. In this setting one may describe a nonequilibrium steady-state (NESS) for which the probability distribution of total entropy production obeys a detailed fluctuation theorem. One can further coarse grain the mesostates into functionally distinct macrostates, equipped under certain conditions with a notion of nonequilibrium entropy production [Seifert (2019)]. Thermodynamic *inference* then refers to extracting information about hidden states from the observable trajectories that would otherwise remain inaccessible, without application of the stochastic thermodynamic framework. Finally, thermodynamic uncertainty relations established within the stochastic thermodynamic framework allow one to put bounds on the energetic cost necessary for accurate timekeeping at the molecular level [Barato and Seifert (2016)].

3.2 Talk 2. Sarah Harvey: An Energy-Accuracy Tradeoff in Nonequilibrium Cellular Sensing

Building on the framework reviewed in Seifert's talk, Sarah Harvey addressed the question of the relationship between entropy production and measurement precision in chemical networks, when the network functions to estimate the concentration of a chemical signal [Harvey et al (2020)]. This question goes back to classical work of [Berg and Purcell (1977)] that was revisited in an ideal observer framework in [Endres and Wingreen (2009)], and studied as an information theoretic problem in [Thomas and Eckford (2016)]. Exploiting the stochastic thermodynamics framework and large deviation theory, Harvey derived two theoretical bounds on the uncertainty of a sensor modeled as a continuous-time Markov process, in different limits of what is observable about the process. The Cramèr-Rao bound for an ideal observer gives $\mathrm{var}(\hat{c})/c^2 \geq 1/\overline{N}$, where $\overline{N}$ is the expected number of binding events in a fixed observation time $T_{\rm obs}$, c is the concentration that is to be estimated, and $\hat{c}$ is the estimate. In contrast, the "coarse-grained" bound for a simple observer gives $\mathrm{var}(\hat{c})/c^2 \geq 8/(T_{\rm obs}\Sigma^\pi + 4\overline{N})$, where Σ^π is the entropy production rate. In particular, Harvey showed there is no advantage to endowing the signal transduction network with additional states beyond "bound" and "unbound".

3.3 Talk 3. Massimiliano Esposito: Thermodynamics of Biochemical Reaction Networks: Information, Accuracy and Speed

Massimiliano Esposito began by reviewing deterministic aspects of open chemical reaction networks (CRNs). As in Seifert's description of a "supersystem" in which a "core system" and a "reservoir" are embedded, an open CRN allows for exchange of energy and matter with its surroundings. Objects used in network analysis of chemical reaction systems such as the stoichiometry matrix naturally partition into components defined by the core/reservoir distinction, leading to versions of the first and second law of thermodynamics adapted to this setting [Rao and Esposito (2016)]. The topology of the CRN and any resulting conservation laws impact the entropy production, thus even for models of *deterministic* chemical reaction systems, thermodynamics and information are fundamentally related to one another. These observations lead to a relation between the relative entropy and the minimum work needed to generate a nonequilibrium distribution, starting from equilibrium (which equals the maximum work that could be extracted from that nonequilibrium distribution as the system approaches equilibrium) [Falasco et al (2018)]. One can analyze open chemical systems as thermodynamic machines and design systems for self-assembly [Penocchio et al (2019)]. This framework allows one to assess the cost, accuracy and speed of various cellular operations, such as energy transduction from molecular motors to metabolism, and the cost of cellular information processing and computation.

Bridging from the deterministic to the stochastic thermodynamic setting, Esposito observed that the underlying structure of thermodynamics carries over unchanged, with thermodynamic entropy becoming the Shannon entropy of the probability of species abundances, and with entropy production satisfying a fluctuation relation [Rao and Esposito (2018)]. In some cases the stochastic and deterministic descriptions are

equivalent, for instance when the CRN is linear, or when a CRN with a network deficiency is at steady state. In general, however, strict equivalence is not satisfied. Finally, in order to rigorously treat energetic and information processing constraints on biological systems beyond the subcellular level, detailed accounting of energetic would be required, which remains a daunting challenge [Esposito (2020)].

3.4 Talk 4. Thomas Ouldridge: Non-Equilibrium Thermodynamics of Catalytic Information Processing

A catalytic information processing system (as introduced in Seifert's talk) is a communication system in which the state of a receiver (e.g. a receptor protein) is "copied" to the state of a readout molecule, without consuming or altering the receiver. Examples include cell surface receptors, but also DNA (the DNA molecule is not consumed or altered in the process of transcription) and RNA (RNA's involvedment in translation may be considered catalytic). As a hallmark of catalytic information processing, the effect of the input persists beyond the timescale of the substrate/catalyst interaction. This extended persistence effect can be exploited for signal amplification, signal splitting, time integration, and modularity. [Ouldridge (2018)]

However, catalytic molecular systems are challenging to design and build. In thermodynamic terms, one considers two distinct macrostates m and m' (as in Seifert's formulation) each containing several microstates $y \in m$, $y' \in m'$. The probability of the microstate $p(m)$, which indicates how far m is removed from equilibrium, is the key quantity to consider, along with the generalized free energy $G[p(m)] = U[p(m)] - TS[p(m)]$. Information transfer from the input signal to the output signal requires occupation of macrostates far from equilibrium. In contrast to recent feats of nanoengineering involving specification of *equilibrium* states (e.g. self-assembly of molecular structures), producing specific non-equilibrium states remains remarkably difficult. As Ouldridge explained: the reason there are very few examples of synthetically engineered catalytic information processing systems (despite the nanoengineering field's track record of success with self-assembly) is that in principle they require strong, selectively-attractive interactions that can be disrupted later, when no longer needed. To address these problems, Ouldridge and colleagues have investigated ways to optimize enzymatic catalysts for rapid turnover of substrates, with low enzyme sequestration; results were published in [Deshpande and Ouldridge (2020)].

3.5 Talk 5. Ilka Bischofs: Information Processing by Bacterial Quorum Sensing Systems

Signal transduction in bacteria provides important examples of communications systems at the level of single cells and, in the case of quorum sensing, populations of cells. Quorum sensing allows the bacteria in a colony to communicate, via secretion and detection of autoinducer molecules, in order to undertake collective actions that individual bacteria could not accomplish alone. Examples include production of bioluminescence that provides a symbiotic advantage to a multicellular host organism such as the angler fish [Nealson and Hastings (1979)], formation of biofilms, induction of virulence factors, initiation of sporulation, production of antibiotics to suppress competing, and many other actions [Miller and Bassler (2001), Mukherjee and Bassler (2019)].

Bischoffs described the commonly observed one-component signaling systems, of which the lac operon is a canonical example, and the less common two-component system, of which histidine kinase signaling or cheA signaling in chemotaxis are well studied examples, provide instances of "catalytic information processing" in the sense discussed by Ouldridge earlier in the workshop. Auto-inducer systems involved in quorum sensing provide another class of examples. Auto-inducer systems are commonly thought to be detecting population density, triggering a population-level response when the density exceeds some threshold. Specific examples of have been reported as early as 1964 in gram-positive pneumococcus bacteria (an activator-inhibitor system regulating competence for genetic transfer [Tomasz and Hotchkiss (1964)]) and 1970 in gram-negative bacteria [Nealson et al. (1970)].

Quorum sensing systems exhibit a variety of network architectures, often (but not always) involving positive feedback loops. Recent work has emphasized the importance of adopting a modular view of quorum sensing systems, representing encoding and decoding as two distinct aspects [Drees et al (2014)]. In the majority of previous work in quorum sensing, bacterial populations are conceived as homogeneous, and synchronized

in their response to changes in population density. In contrast, Bischof's recent work has emphasized heterogeneous and heterochronous aspects of quorum-sensing populations and responses [Bettenworth et al (2019)] as well as a novel pump-probe model of ratiometric population sensing [Babel et al (2020)].

4 Scientific Progress Made

Given the reduced scope of the workshop, and the lack of opportunities for informal face-to-face conversations (on the trail up Tunnel Mountain, for example) the organizers did not expect to report immediate significant scientific advances. However, as one of the few conferences devoted solely to exploring the implications of stochastic thermodynamics in biological systems, the workshop served as a meeting ground for researchers in this area, and started new conversations. Several participants reported following up with new contacts initiated through the workshop. As one participant (V. Klika) wrote "taking part in the workshop has finally pushed me to carry out an idea I had for many years now – the effect of kinetic energy density on reaction-diffusion model when derived via non-equilibrium thermodynamics."

5 Outcome of the Meeting

The original organizers of 20w5074 planned to coordinate a special issue of the journal *Biological Cybernetics* on the topic of "Information Theory and Thermodynamics in Biology" in conjunction with the BIRS workshop.

Following the onset of the coronavirus pandemic, two of the original three co-organizers resigned from organization of the conference, and were replaced with new co-organizers. The workshop was reorganized on a smaller scale with five talks and a poster session spread over three days. The planned special issue attracted only four submissions; as of the submission of this report one paper was published as part of a regular issue of the journal [Deshpande and Ouldridge (2020)], one paper was rejected, one is being revised by its authors, and one remains under review.

On the positive side, the online format allowed for greater participation in the workshop than the in-person format would have allowed. Over 60 participants attended each session of the online workshop. One of the speakers (Seifert, a leading figure in the field) had previously declined to participate because of being unable to travel, but was able to give the keynote talk remotely.

References

[Babel et al (2020)] H. Babel, P. Naranjo-Meneses, S. Trauth, S. Schulmeister, G. Malengo, V. Sourjik and I.B. Bischofs, Ratiometric population sensing by a pump-probe signaling system in Bacillus subtilis. *Nature communications* **11**.1 (2020), 1–13.

[Barato and Seifert (2016)] A.C. Barato and U. Seifert, Cost and precision of Brownian clocks, *Physical Review X* **6**.4 (2016), 041053.

[Berg and Purcell (1977)] H.C. Berg and E.M. Purcell, Physics of chemoreception, *Biophysical journal* **20**.2, (1977) 193–219.

[Bettenworth et al (2019)] V. Bettenworth, B. Steinfeld, H. Duin, K. Petersen, W.R. Streit, I. Bischofs and A. Becker, Phenotypic heterogeneity in bacterial quorum sensing systems, *Journal of molecular biology*, **431**.23 (2019), 4530–4546.

[Collin et al (2005)] D. Collin, F. Ritort, C. Jarzynski, S.B. Smith, I. Tinoco and C. Bustamante, Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies, *Nature* **437**.7056 (2005), 231–234.

[Deshpande and Ouldridge (2020)] A. Deshpande and T.E. Ouldridge, Optimizing enzymatic catalysts for rapid turnover of substrates with low enzyme sequestration, *Biological Cybernetics* **114**.6 (2020), 653–668.

[Drees et al (2014)] B. Drees, M. Reiger, K. Jung and I.B. Bischofs, A modular view of the diversity of cell-density-encoding schemes in bacterial quorum-sensing systems, *Biophysical journal* **107**.1 (2014), 266–277.

[Endres and Wingreen (2009)] R.G. Endres and N.S. Wingreen, Maximum likelihood and the single receptor, *Physical review letters* **103**.15 (2009), 158101.

[Esposito (2020)] M. Esposito, Open questions on nonequilibrium thermodynamics of chemical reaction networks, *Communications Chemistry* **3**.1 (2020), 1–3.

[Falasco et al (2018)] G. Falasco, R. Rao and M. Esposito, Information thermodynamics of turing patterns, *Physical Review Letters* **121**.10 (2018), 108301.

[Harvey et al (2020)] S. Harvey, S. Lahiri and S. Ganguli, An Energy-Accuracy Tradeoff for Nonequilibrium Receptors, *Bulletin of the American Physical Society* **65** (2020).

[Horowitz and Esposito (2014)] J.M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, *Physical Review X* **4**.3 (2014), 031015.

[Miller and Bassler (2001)] M.B. Miller and B.L. Bassler, Quorum sensing in bacteria, *Annual Reviews in Microbiology*, **55**.1 (2001), 165–199.

[Mukherjee and Bassler (2019)] S. Mukherjee and B.L. Bassler, Bacterial quorum sensing in complex and dynamically changing environments, *Nature Reviews Microbiology* **17**.6 (2019), 371–382.

[Nealson et al. (1970)] K.H. Nealson, T. Platt and J. W. Hastings, The cellular control of the synthesis and activity of the bacterial luminescent system, *J. Bacteriol.* **104** (1970), 313–322.

[Nealson and Hastings (1979)] K.H. Nealson and J.W. Hastings, Bacterial bioluminescence: its control and ecological significance, *Microbiological reviews* **43**.4 (1979), 496–518.

[Ouldridge (2018)] T.E. Ouldridge, The importance of thermodynamics for molecular systems, and the importance of molecular systems for thermodynamics, *Natural computing* **17**.1 (2018), 3–29.

[Penocchio et al (2019)] E. Penocchio, R. Rao, and M. Esposito, Thermodynamic efficiency in dissipative chemistry, *Nature communications* **10**.1 (2019), 1–5.

[Rao and Esposito (2016)] R. Rao and M. Esposito, Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics, *Physical Review X* **6**.4 (2016), 041064.

[Rao and Esposito (2018)] R, Rao, and M. Esposito, Conservation laws and work fluctuation relations in chemical reaction networks, *The Journal of Chemical Physics* **149**.24 (2018), 245101.

[Seifert (2008)] U. Seifert, Stochastic thermodynamics: principles and perspectives, *The European Physical Journal B* **64**.3-4 (2008), 423–431.

[Seifert (2012)] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Reports on progress in physics* **75**.12 (2012).

[Seifert (2019)] U. Seifert, From stochastic thermodynamics to thermodynamic inference, *Annual Review of Condensed Matter Physics* **10** (2019), 171–192.

[Thomas and Eckford (2016)] P.J. Thomas and A.W. Eckford, Capacity of a simple intercellular signal transduction channel, *IEEE Transactions on Information Theory* **62**.12 (2016), 7358–7382.

[Tomasz and Hotchkiss (1964)] A. Tomasz and R.D. Hotchkiss, Regulation of the transformability of pneumococcal cultures by macromolecular cell products, *Proceedings of the National Academy of Sciences of the United States of America* **51**.3 (1964), 480–487.