Application of Finite-Time and Control Thermodynamics to Biological Processes at Multiple Scales

Roach, Ty N. F.; Salamon, Peter; Nulton, James; Andresen, Bjarne; Felts, Ben; Haas, Andreas; Calhoun, Sandi; Robinett, Nathan; Rohwer, Forest

Published in:
Journal of Non-Equilibrium Thermodynamics

DOI:
10.1515/jnet-2018-0008

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Review

Ty N. F. Roach*, Peter Salamon, James Nulton, Bjarne Andresen, Ben Felts, Andreas Haas, Sandi Calhoun, Nathan Robinett, and Forest Rohwer

Application of Finite-Time and Control Thermodynamics to Biological Processes at Multiple Scales

https://doi.org/10.1515/jnet-2018-0008
Received March 5, 2018; revised April 27, 2018; accepted May 2, 2018

Abstract: An overall synthesis of biology and non-equilibrium thermodynamics remains a challenge at the interface between the physical and life sciences. Herein, theorems from finite-time and control thermodynamics are applied to biological processes to indicate which biological strategies will succeed over different time scales. In general, living systems maximize power at the expense of efficiency during the early stages of their development while proceeding at slower rates to maximize efficiency over longer time scales. The exact combination of yield and power depends upon the constraints on the system, the degrees of freedom in question, and the time scales of the processes. It is emphasized that biological processes are not driven by entropy production but, rather, by informed exergy flow. The entropy production is the generalized friction that is minimized insofar as the constraints allow. Theorems concerning thermodynamic path length and entropy production show that there is a direct tradeoff between the efficiency of a process and the process rate. To quantify this tradeoff, the concepts of compensated heat and waste heat are introduced. Compensated heat is the exergy dissipated, which is necessary for a process to satisfy constraints. Conversely, waste heat is exergy that is dissipated as heat, but does not provide a compensatory increase in rate or other improvement. We hypothesize that it is waste heat that is minimized through natural selection. This can be seen in the strategies employed at several temporal and spatial scales, including organismal development, ecological succession, and long-term evolution. Better understanding the roles of compensated heat and waste heat in biological processes will provide novel insight into the underlying thermodynamic mechanisms involved in metabolism, ecology, and evolution.

Keywords: evolution, ecology, succession, information, entropy, exergy, minimum dissipation, maximum power, compensated heat, waste heat

1 Introduction

Many scientists have contributed to understanding the underlying thermodynamics of biological processes. One of the most notable attempts was Erwin Schrödinger’s book What is Life [1]. The purpose of the current
paper is to provide an updated perspective on the role of thermodynamics in biological systems. Many such updates appear with some regularity (see for example Murphy and O’Neil’s *What is Life: The next 50 Years* [2]). The present update differs in two respects. First, it reformulates evolution and ecology in terms of informed *exergy* flows and controlled *spontaneity*, shifting the focus away from *entropy* production and dissipation (Box 1 and Table 1). Second, we explicitly examine the implications of finite-time and control thermodynamics for biological processes. It is our hope that this paper will be of value to biologists and physical scientists alike. To this end, both equations and biological jargon have been kept to a minimum. There are also a number of text boxes to afford deeper explanations of concepts that may need greater explanation than allowed for in the main text, as well as a table of terms (Table 1).

**Box 1 (Exergy and entropy).**

The *exergy* (A) of a thermodynamic system is the work that could be extracted in a process that reversibly brings the system to equilibrium with its environment [3]. For processes at fixed environmental temperature and pressure, the change in exergy is just the change in free energy of the system. In any spontaneous process, exergy is dissipated, and available work is irretrievably lost. This dissipated exergy appears as heat.

*Entropy* (S) is a thermodynamic state function whose change is equal to the heat received by the system divided by its temperature. Entropy is also the logarithm of the number of microstates (W) consistent with the information (I) about the system \( S = k_B \ln W \). In statistical terms this has been expressed as the well-known Shannon formula \( S = -k_B \sum p_i \ln p_i \). This statistical treatment of entropy has led to entropy also being defined as missing information \( S = -I \). Here, \( k_B \) is the Boltzmann constant \( 1.38064852 \times 10^{-23} \text{ J} \cdot \text{K}^{-1} \) and \( p_i \) is the probability of state \( i \).

## 2 Thermodynamic perspectives

### 2.1 Thermodynamics of spontaneous exergy flow

One perspective of thermodynamics is the quantitative characterization of *spontaneity* of a flow. The motivation for this notion began with a water wheel, where the spontaneity of water flowing is manifest as mass moving toward a lower gravitational potential. The spontaneity of this flow is the maximum work, which equals the mass flow times the difference in potential, so we have

\[
\text{Spontaneity} = \text{Flow} \times \text{Potential difference}.
\]

This point of view was instrumental to Carnot’s recognition that the work produced by a steam engine is driven by heat flowing to a lower temperature [4, 5]. He further realized that just as for the water wheel, the most work obtainable from a heat engine is the least work required to raise the heat back to the higher temperature. This enabled the creation of a temperature scale in which the temperatures can act as the potential in eq. (1).

Gibbs [6] extended Carnot’s realization to chemical phenomena, again exploiting the maximum work due to the spontaneity of chemical reactions to define chemical potentials such that the maximum work is the stoichiometric combination of the chemical potential differences (the affinity) times the mass that reacted (flowed). Following this point of view, we define the spontaneity of the flow of the \( i \)-th extensity \( x_i \) through the intensity difference \( y_i - y_i^e \) to be

\[
\text{Spontaneity}_i = \int (y_i - y_i^e) \, dx_i.
\]

The spontaneity of one flow can then be coupled to other flows to drive these flows in a specific direction. The sum of spontaneities of all the extensities is guaranteed to be non-negative for all real processes by the second law. Thus, the flow of extensities with negative spontaneity can be “driven” by coupling these to flows with positive spontaneity.
Table 1: Table of terms. All terms are italicized the first time they appear in the text. The table is ordered alphabetically. Symbols used in equations are shown in parentheses.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autopoiesis</td>
<td>A system’s property of being able to maintain and reproduce itself</td>
</tr>
<tr>
<td>Compensated heat (Qc)</td>
<td>The exergy dissipated as heat which is necessary to satisfy constraints</td>
</tr>
<tr>
<td>Constraint</td>
<td>A conditional requirement of the feasible set of solutions to an optimization problem</td>
</tr>
<tr>
<td>Degree of freedom</td>
<td>An independent coordinate, describing the state of a system, along which change can occur</td>
</tr>
<tr>
<td>Dissipation</td>
<td>The degradation of energy available to do work</td>
</tr>
<tr>
<td>Dissipative structure</td>
<td>An open non-equilibrium thermodynamic system that generates order spontaneously by directing streams of exergy</td>
</tr>
<tr>
<td>Emergy</td>
<td>The amount of exergy that is either directly or indirectly required to drive a process</td>
</tr>
<tr>
<td>Entropy (S)</td>
<td>Entropy is a thermodynamic state function whose change is equal to the heat received by the system divided by its temperature. See also Box 1.</td>
</tr>
<tr>
<td>Exergy</td>
<td>The work that could be extracted in a process that reversibly brings the system to equilibrium with the environment. See also Box 1.</td>
</tr>
<tr>
<td>Far-from-equilibrium system</td>
<td>Systems subjected to flows of matter and/or energy, in the non-linear regime</td>
</tr>
<tr>
<td>Fitness</td>
<td>The ability of an entity to survive and reproduce. Fitness is the quantitative parametrization of selection in evolutionary biology</td>
</tr>
<tr>
<td>Functional control region</td>
<td>The region between minimum dissipation operation and maximum power operation</td>
</tr>
<tr>
<td>Heterotroph</td>
<td>An organism that cannot fix carbon from an inorganic source, so it must consume complex organic molecules that have been previously anabolized by another organism</td>
</tr>
<tr>
<td>Informed</td>
<td>Structured by the passage of information from a previous time</td>
</tr>
<tr>
<td>Near-equilibrium systems</td>
<td>Systems in which flows are linear in the forces</td>
</tr>
<tr>
<td>Ontogeny</td>
<td>The development of an organism from the time of its inception (e. g., fertilization) to its mature form</td>
</tr>
<tr>
<td>Orientor</td>
<td>A variable that describes where a system tends move in state space</td>
</tr>
<tr>
<td>Pareto optimum</td>
<td>A set of states which optimizes a combination of two or more objectives (e. g., maximum power and minimum dissipation) where an improvement in one objective cannot occur without the worsening of another</td>
</tr>
<tr>
<td>Power</td>
<td>The rate at which work is done</td>
</tr>
<tr>
<td>Selection</td>
<td>The differential reproduction and survival of a particular type. Selection is the key mechanism behind evolutionary change</td>
</tr>
<tr>
<td>Spontaneity</td>
<td>The maximum work obtainable from a reversible flow of an extensity</td>
</tr>
<tr>
<td>Succession</td>
<td>The process of sequential change in the structure of an ecological community as the ecosystem matures</td>
</tr>
<tr>
<td>Unit of selection</td>
<td>Any entity in the hierarchy of biological organization that is subject to natural selection and subsequent change over time</td>
</tr>
<tr>
<td>Waste heat (Q_w)</td>
<td>The heat dissipated above and beyond the minimum necessary to achieve the specified rate of the process subject to constraints</td>
</tr>
</tbody>
</table>

More generally, the focus here is about informed exergy flow. That is, the flow of exergy through pathways that have been constrained by the passage of information from a previous time. Living systems direct flows of exergy from sources, such as the sun or chemical bonds, through informed pathways to achieve homeostatic and environmental control, and ultimately to reproduce. We emphasize that these pathways are informed to highlight that living systems can extract work from an exergy source only if the exergy flows through specific pathways which are informed by the passage of information, for example, the information stored in the DNA molecules that are passed from parents to progeny. The ability of biological systems to compete for exergy flows and efficiently direct them through informed pathways to extract work is the basis for what defines how well a living entity will survive and reproduce. This differential ability to direct flows of exergy through informed pathways to reproduce is ultimately what determines fitness and long-term evolutionary trajectories.

2.2 Thermodynamics of dissipation

The quantitative thermodynamic measure of spontaneity of a flow is the amount of work the flow could do if it proceeded reversibly. However, real spontaneous processes are irreversible due to dissipation of exergy
resulting in entropy production. Thus, real processes obey rather a balance condition involving exergy where

\[ \text{Exergy}_{\text{in}} - \text{Exergy}_{\text{out}} = \Delta \text{Exergy}_{\text{stored}} + \text{Exergy}_{\text{dissipated}}, \]  

(3)

where the first term on the right includes exergy stored in bonds as well as mechanical energy.

In energetic terms, dissipation is measured by the amount of lost work; in entropic terms it is measured by the amount of entropy produced. The lost work shows up as heat according to the Gouy–Stodola theorem, which states that the reduction in exergy available to do work is proportional to the amount of entropy, \( S \), produced [7, 8, 9, 10]. That is,

\[ \text{Exergy}_{\text{dissipated}} = -T_0 S_{\text{produced}}, \]  

(4)

where \( T_0 \) is the temperature of the environment. Thus, either dissipated work or entropy produced can quantify dissipation.\(^1\)

Dissipation happens when there is a flow in response to a force without full compensation in work stored. One example of this is when one kilogram of water moves down one meter without doing the maximum 9.8 joules of work \((m \cdot g \cdot h)\). The difference between the work done and this 9.8 joules is the dissipation. Lost work appears as heat and equals the temperature of the environment times the entropy produced (see eq. (4)).

Modern fluctuation theory [11, 12, 13] has recently provided us with an intriguing expression for the entropy production of a process given its probability of proceeding forwards, \( P_f \), and backwards, \( P_r \), as

\[ S_{\text{produced}} = k_B \ln \left( \frac{P_f}{P_r} \right), \]  

(5)

where \( k_B \) is Boltzmann’s constant. One interpretation is that the more spontaneously (unidirectionally) a step needs to proceed, the more entropy production must occur. This fact and its various phenomenological counterparts have led many authors to assert that the entropy production “drove” the process [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. However, entropy production does not drive any process; it is a consequence of the rate of operation. The flow of exergy through informed pathways is what drives living systems to do work and store exergy. The dissipation is the generalized friction in these processes that results due to the constraints of finite time. Processes get selected to minimize the friction (the lost exergy, the wasted work, the entropy produced) while getting the job done with the time and resources available.

In terms of eq. (5), consider bringing together two macromolecular reactants in a way that significantly changes the odds that two reaction centers align. Although this may incur a cost, a molecular motor can make this happen by bringing the reactants together in just the right way. Having such a molecular motor dramatically changes the odds ratio in eq. (5).

Flows of exergy during spontaneous reactions result in entropy production. This entropy production, like friction in mechanics, is a necessary byproduct due to imperfect technology. The technology makes a huge difference, and for living systems this “technology” comes in the form of informed exergy flow pathways such as molecular motors and carefully choreographed reaction chains informed by DNA [23]. Note that exergy not wasted in one step is available for use in other coupled processes.

Many authors, starting with Lotka in 1922 [24, 25], have said much of the above: that exergy flows drive and hence makes possible the control of events in biological systems. Recognizing that dissipation is not what drives biological processes, but is rather the undesirable part of what the technology allows, is however an important new twist [26].

### 3 Thermodynamic states of biological systems

Both thermodynamics and biology frequently use the notion of equilibrium states. It is important to realize that these notions of equilibrium depend on the timescale at which one is looking as well as on the de-
Figure 1: Dynamics across timescales. In any experiment, an observer is limited to certain degrees of freedom that are observably out of equilibrium. At different observational time scales, different degrees of freedom will have observable dynamics. Other degrees of freedom will be too fast. That is, relaxation time is too fast to observe dynamics in these degrees of freedom and thus they appear equilibrated. Still other degrees of freedom will be too slow. This means that relaxation time is too slow to observe dynamics in these degrees of freedom causing them to appear stationary or non-existent. Thus, only the degrees of freedom with relaxation times on the same scale as the observation time will have observable dynamics.

degrees of freedom of interest. Every process has a timescale of relevance. For an organism's processes this timescale could be milliseconds to years, for an ecosystem hours to decades. Processes faster than this relevant timescale will be fully equilibrated and thus not observable. Processes that are slower than this will appear as stationary since nothing changes within the observed time window. Thus, being observably out of equilibrium and the rate at which the system approaches equilibrium depend critically on the timescale (Figure 1) [27, 28, 29].

A concept that is central to this relation between equilibrium and timescales is the notion of degrees of freedom. The independent coordinates, describing the state of a system, along which change can occur are the system's degrees of freedom. Real biological systems contain many different degrees of freedom. On the other hand, for a diatomic molecule, there are relatively few. There are translational degrees of freedom of the molecule as a whole along the X, Y, and Z directions. There are also two rotational degrees of freedom and one vibrational degree of freedom. For more complicated systems, inclusion of diffusion, charge build-up, temperature gradients, etc. may be relevant degrees of freedom. Each of these may be fully equilibrated and may also be equilibrated with some others while at the same time totally out of equilibrium with the remaining degrees of freedom. For example, translational motion is what is normally associated with a sense of temperature, but rotation and vibration will often have their own temperatures of quite different values. The fact that this occurs is a sign that translational motion does not couple strongly with rotation and vibration. In that case, rotation and vibration may be disregarded when describing a chemical reaction at (translational) temperature $T$. Only those degrees of freedom which change significantly on our timescale need to be described in detail; all the other degrees of freedom can be treated parametrically on that time scale.

As a simple illustration of the different dynamic behaviors of degrees of freedom, consider the temperature and pressure in a sealed room, both being different from the outside temperature and pressure. Then slightly open one of the windows. Pressure will equilibrate almost instantaneously while temperature will take quite some time (i.e., those two equilibration processes have distinctly different time scales). Add the CO$_2$ concentration as an additional degree of freedom, and you have a still more complicated equilibration process. Hence, the notion of how close a system is to an equilibrium state depends both on the relevant timescale and the degrees of freedom being observed (Figure 1).

Irreversible thermodynamics often separates processes as being near equilibrium and far from equilibrium. The former concerns systems where the flows are in the linear regime (i.e., flows are proportional to the corresponding forces). In the latter category forces are higher and the flows are non-linear. Thus, what is considered is the rates of approach to equilibrium, not the actual forces.
3.1 Non-equilibrium thermodynamic states and dissipative structures

Lotka pointed out that although living systems may appear to be unchanging and near equilibrium, they might in fact be in a non-equilibrium steady state “feeding” off an exergy gradient to stay some distance away from equilibrium. Onsager [30, 31, 32] provided one of the first complete theoretical discussions of non-equilibrium steady-state dynamics and demonstrated that near-equilibrium steady states could persist as long as there was an available exergy stream. He further showed that these near-equilibrium systems with an influx of exergy tend toward states of minimum specific dissipation rates [30, 31]. These states consist of structured systems that produce the minimal entropy compatible with the given constraints.

Box 2 (Dissipative Structures).

**Dissipative structures** are open non-equilibrium thermodynamic systems that generate order spontaneously by directing streams of exergy. A classic physical example of a self-organizing dissipative system is the Benard cell [33, 34, 35]. In this example, a temperature gradient is applied to a fluid in a gravitational field, which causes an instability in the system leading to self-organization in the form of convective cells. The Benard cell is an example of order increasing within a system due to exergy flow and the concomitant production and export of entropy across the boundaries of a dissipative system. Another example of self-organizing dissipative systems are Taylor vortices. Taylor vortices can be seen in nature as hurricanes, tornados, and whirlpools [35, 36]. This self-organization initially increases the entropy production rate, followed by a decrease in the entropy production rate toward a minimum at steady state [68].

Figure 2: Schema of dissipative structures at multiple biological scales. A: Schematic representation of a dissipative structure that arises when systems are situated in an exergy gradient. These systems utilize low entropy exergy to do work and store exergy in forms such as chemical bonds. Dissipative structures can be coupled to other systems (green arrows) or perturbed by other systems (orange arrow). B: Examples of dissipative structures at multiple biological scales from the cellular systems of microorganisms being perturbed by phages on the left to organisms such as plants and animals interacting with environmental parameters to ecosystems such as coral reefs being perturbed by human impact shown on the far right.
Following from Onsager’s work, Prigogine [37, 38] and the Brussels School [39, 40, 41] described the dynamics of far-from-equilibrium systems and characterized a thermodynamic construct suitable for describing living systems from cells to ecosystems. They termed these systems dissipative structures [37] (Box 2, Figure 2A). Dissipative structures form when systems are situated in a gradient, be it chemical, thermal, or gravitational, that induces exergy flow through the system [42, 43, 44]. These flows can power the export of entropy generated in the system into the surrounding environment. Thus, these systems maintain their own entropy below maximal, allowing for the maintenance or creation of order and organization (Figure 2A) [15, 16, 45, 46, 47, 48, 49, 50, 51].

3.2 Life-like features of dissipative structures

In addition to being open, irreversible systems far from equilibrium, some dissipative systems are also autocatalytic. Well-known examples of such systems are the Belousov–Zhabotinsky reaction [52] and the self-replicating hypercycles of Eigen and Schuster [53, 54, 55]. Prigogine and the Brussels school [38, 39, 40, 42] showed that autocatalytic systems are able to cycle materials [48] and form feedback loops [56]. When autocatalytic cycles give rise to feedback loops, the first element of the loop (the input) will be affected by the last element (the output), resulting in self-organization of the entire system [57, 58, 59]. That is, the feedback loop, driven by an external flow of exergy, forms an autonomous network, in which one element of the network can exert control on the rest of the cycle. Fluctuations in these dissipative systems can yield instabilities that can cause spontaneous hierarchical bifurcations (e.g., Turing bifurcations) [41, 42, 44, 60, 61, 62]. These bifurcations can then give rise to variation, which can undergo natural selection (Table 1). They can also create nested hierarchies of dissipative systems within dissipative systems [62, 63, 64]. These hierarchies then lead to networks of interactions that will determine the optimal operation of each subsystem. In this sense, biological systems across scales from cells, to organisms, to ecosystems can be seen as networks of dissipative structures (Figure 1B). Dissipative structures can be autocatalytic [37], self-organized [56, 58, 65] systems capable of settling to multiple far-from-equilibrium steady states [66] and prone to hierarchical bifurcations [61, 62]. Thus, it is generally accepted that dissipative structures (Figure 2A) are the closest thermodynamic representations of biology at all scales (Figure 2B).

4 Finite-time thermodynamics

4.1 Tradeoffs between power and yield

The tradeoff between power and efficiency has been the major theme of finite-time thermodynamics (FTT) [67, 68, 69, 70]. FTT studies concerning various types of heat engines reveal a rich tapestry of optimal operations for various possible constraints, objectives, and mechanisms [71, 72, 73, 74, 75, 76, 77]. One feature however is invariant. There is an interval of optimal operation between maximum power and minimum dissipation. Operating outside of this interval is usually possible at both ends but entails waste of at least one of the objectives as compared to what is possible within this interval, where better results can be achieved with less resources. Similar results are found for chemical reaction systems [78, 79, 80, 81]. This area, between maximum power and minimum dissipation, where a combination of time and efficiency of resource use can be optimized, is termed the Functional Control Region (FCR) (the green region in Figure 3).

Figure 3 summarizes this in the context of heat engine optimizations. Note that the ordinate, power, being equal to output per time, inherently takes the rate of the process into account. The extremes of optimal operation are minimum dissipation and maximum power [82]. Other objectives are combinations of these [83, 84]. This tradeoff was originally cast in the form of the relative prices of products vs. reactants [82]. Panel A of Figure 3 is the generic tradeoff curve between power and yield (efficiency) for the idealized case of an endoreversible process [85] without idling costs (i.e., no losses associated with doing nothing). The maximum yield is achieved only for reversible operations where everything proceeds infinitesimally slowly and
Figure 3: Typical curves of power and yield. A: The power vs. yield curve for an endoreversible process without idling costs. B: The yield vs. power curve for an endoreversible process with idling costs where dissipation is present even with no power output. Point i on the graphs represents maximum power operation, and point ii represents minimum dissipation operation. The functional control region (FCR), shaded in green, where the slope is less than or equal to zero, is the only portion of the curves where a combination of power and yield is optimized.

thus produces zero power (point ii in Figure 3A). Alternatively, in the limit of very rapid operation, all input is lost to dissipation and thus produces no yield and also no power. Panel B shows the corresponding tradeoff between power and yield when dissipation is present even when no output is produced [86, 87, 88]. In this more realistic case, there is not only an upper bound on the power produced, but also on the efficiency or yield of the process. Only under very special conditions do these two points coincide [75]. Generally, there is the functional control region of tradeoff, marked in green in Figure 3, within which other objectives determine the optimal mode of operation.

4.2 Minimum dissipation, maximum power, and thermodynamic length

The extremes of maximum power and minimum dissipation both introduce interesting conditions of relevance for biology. At the maximum power end there is the universal theorem that the yield is 50% of the reversible yield to lowest order in the forces [89, 90]. In biology, the fact that maximal rate operation foregoes about 50% of the yield is significant. Nonetheless, this theorem from FTT does not seem to be well known. For example, in Karr et al.’s [91] complete simulation of a *Mycobacterium genitalium*, they managed to account for each molecule, with the exception of a couple of water molecules, yet 44% of the exergy used as measured by ATPs produced was unaccounted for. This discrepancy may be explained by the all-out reproduction process being near the universal 50% efficiency at maximum power and the fact that their simulated cell is reproducing as fast as it can. Biological systems working at maximum power are also significant early in colonization events, when all-out growth is called for [92] and when the environment provides an abundance of exergy streams. This is elaborated in Section 5 below.

FTT results concerning minimum dissipation point to a strategy involving a sequence of stages through which the system should pass to minimize the dissipation. The more stages a relaxation process uses, the less dissipation, provided that the stages are appropriately distributed. For a $K$-relaxation approximation to a quasistatic process, the minimum dissipation is

$$\Delta S_{\text{min}} = \frac{L^2}{2K},$$

where $L$ is the total thermodynamic length [93, 94, 95] and $K$ is the number of stages in the process. This equation shows that decreasing the dissipation of a process can be accomplished by either increasing the number of stages $K$, or by shortening the thermodynamic length $L$ between the initial and final states [95]. Note that although adding a step anywhere in the process will reduce dissipation, the steps must be equidistant in the thermodynamic length $L$ to produce the minimum dissipation [93, 94]. Achieving this condition
usually requires additional controls on the operation of the system [97]. This theorem leads us to hypothesize that evolution selects for equal spacing of thermodynamic steps in biological processes.

One illustrative biological example is the electron transport chain (ETC) in mitochondria [81]. If an additional molecule were to be added to the ETC, evolutionary processes should select for the redox potential of the electron in the molecule such that the thermodynamic length should be equal between two existing redox states in the ETC. By utilizing this theorem, one cannot predict which new molecule will arise in the ETC. However, given a choice of molecules from several, one can predict which of the molecules would provide the greatest fitness advantage, and thus, which molecule would be the most likely to be selected for and eventually fixed within a population. This theorem also allows for thermodynamic-based predictions at higher biological levels, such as where new organisms might benefit most in trophic chains, or which new niche space would be the most beneficial to occupy in order to receive the greatest bioenergetic benefit.

Frederiksen and Andresen [81] used eq. (6) to calculate the entropy production associated with the cytochrome chain in human mitochondria, in which the chemical potential of an electron is lowered by sequential stages to extract work and produce, ultimately, ATP molecules. They demonstrated that the spacing of electron carrier molecules in actual mitochondria serves to produce close to the minimum entropy for a \( K \)-step process of a given total path length. These calculations suggest that this strategy of equally spacing each stage to minimize entropy production is being utilized by the mitochondrial ETC.

4.3 Accounting for dissipation

One important lesson from FTT is the ability to account for dissipation. That is, how much dissipation is due to which factors and, most importantly, how much of the dissipated heat is waste. For example, when sunlight heats a dark rock, almost all of the light’s exergy is dissipated completely without any other effect. While some meager fraction of the exergy in the warm rock indeed can go on to contribute to meteorological or erosion effects, the exergy carried by the sunlight effectively at 6000 degrees was degraded to exergy at a few hundred degrees with no other effect. Exergetically, this is a complete waste. It is akin to a waterfall without a water wheel. If the same amount of sunlight is absorbed by plants and the exergy is partially stored as biomass, the dissipation is smaller. Eventually, of course, that stored exergy will end up as heat as well, but it will have accomplished desired biological objectives along the way and can be sequestered over long time periods as seen in large coal and petroleum deposits. It must be emphasized here that what is considered a “desired” effect depends entirely on the controller. For evolution, the controller is natural selection, in the sense that it is natural selection that dictates the evolutionary constraints on an entity and causes it to evolve.

Knowing the controller, however, allows us to minimize the dissipation subject to the constraints chosen by our agent. This minimum dissipation is the compensated heat \( (Q_C) \). That is the exergy dissipated as heat that is compensated for by satisfying the constraints of the controller. Any remaining dissipation above this minimum is waste \( (Q_W) \). Formally, waste is defined as any suboptimality in dissipation.

As an example, consider a heat engine of the type shown in Figure 3A, operating subject to the constraint of a given power \( P = P_0 \). One possible operation is shown as point i in Figure 4. The figure shows how actual yield is the reversible yield minus the compensated heat and waste. Although the process at point i is operating at the necessary rate (i.e., \( P = P_0 \)), there is a large amount of waste, as the yield is lower than it could be if the process were operating at point iii.

Now, consider the maximum power operation point ii in Figure 4. This operation has no waste – the yield is reduced by the dissipation, but this was a necessary cost of increasing power.

For both examples and quite generally,

\[
W_{\text{real}} = W_{\text{reversible}} - Q_{\text{compensated}} - Q_{\text{waste}},
\]

where \( W \) and \( Q \) are work and heat, respectively. The sum of the last two terms is the uncompensated heat of De Donder [98].
Box 3 (Compensated heat and waste heat).

Compensated heat is the exergy dissipated as heat which is necessary to provide a specified rate of a process subject to constraints. Waste heat is the heat dissipated in excess of the minimum necessary to achieve the specified rate of the process subject to constraints. The sum of compensated heat and waste equals the total dissipation. The decomposition depends on the specified constraints. Note that the environment here includes a specification for the available technology, so the optimization problem is well-defined.

Figure 4: Schematic representation of compensated heat ($Q_C$) and waste heat ($Q_W$) as a function of power and yield. The horizontal dashed line is a constraint on the process that the power must be equal to $P_0$. Both points i and iii satisfy the constraint that $P = P_0$. However, the process proceeding at point i has a lower yield and more exergy is dissipated as waste, whereas at point iii there is no waste, as all of the exergy dissipated as heat is compensated by meeting the constraint that $P = P_0$.

Such analysis accounts for the dissipation as having been spent to satisfy a constraint. For biological systems, though, the cost is only partially due to satisfying rate constraints. They also incur costs due to resource scarcity and other environmental factors, as well as costs associated with spontaneity amplification of the sort previously discussed in connection with eq. (5). Molecular processes in living cells must take place with high yield and high fidelity. This forces many sizeable dissipation terms of the form $k_B \ln(P_f/P_r)$. Other compensated dissipation costs may come from unknown constraints. For example, organisms need the ability to operate in more than one mode to achieve different goals such as when otherwise efficient muscles shiver to generate heat [99].

Accounting for waste needs a detailed specification of the technology used in the optimization problem. One concrete physical example is the losses associated with separating a mixture of two liquids into its components. Traditional distillation column technology can never achieve a thermal efficiency greater than 69% simply by its construction [100]. Switching to diabatic column technology eliminates this inefficiency. In biological systems, the specification of the technology is the set of available informed exergy flow pathways, such as enzymes and molecular motors.

5 Thermodynamics across multiple biological scales

Non-equilibrium thermodynamics highlights the fact that there are specific generalities concerning entropy production in thermodynamic systems at all scales (Figure 2). FTT takes this one step further by taking time and rate into account and demonstrating that there is an explicit tradeoff between power and yield (Figures 3 and 4). This tradeoff provides an interesting optimization problem, where natural selection serves to optimize some combination of yield and power (i.e., a Pareto optimum). Because of the constraints and selective pressures of FTT, we find that there is a general power / yield tradeoff principle whereby biological systems maximize power at the expense of yield over short time scales and when resources are plentiful, while proceeding at slower rates to maximize efficiency over longer time scales or when resources are scarce.
This principle leads to several hypotheses and predictions about how biological systems will behave over particular time scales and exergy regimes. In the following sections we highlight several empirical studies which serve as illustrative examples of how this principle explains the behavior of biological processes across multiple scales (Figures 1 and 2), including ontogeny, succession, and evolution.

5.1 Ontogeny and aging

Prigogine and Wiame [101] hypothesized that the specific entropy production rate (i.e., entropy production rate per unit biomass) decreases to a stable minimum during the development of biological systems. Zotin, Lamprecht, and Zotina [102, 103, 104, 105] tested the predictions of the Prigogine–Wiame hypothesis [101], demonstrating empirically that both ontogeny and development are characterized by a steady decrease in the specific entropy production rate, although often preceded by an initial increase in specific entropy production rate.

This hypothesis has been further empirically supported by data on trout [106], frog, and chicken [107, 108] eggs during ontogeny and during post-embryonic development on bivalves [109, 110], gastropods, insects [111, 112], birds [113, 114], amphibians [115, 116], and even humans [96, 117, 118]. These data support the prediction that the specific dissipation rate eventually decreases throughout the development and growth of organisms despite sometimes increasing initially.

5.2 Ecology and succession

FTT can also be applied to the process of sequential change in the structure of ecological communities, known as succession (Table 1). As an example, studies on marine microbes demonstrate that microbial communities in marine systems undergo succession to optimize the tradeoff between yield and power in a predictable manner [119]. The heterotrophic portion of these communities uses organic carbon as its exergy source. When sea waters have greater bioavailable exergy stored in the form of these organic carbons the heterotrophic communities switch their central carbon catabolism toward alternative glycolytic pathways with fewer thermodynamic steps (i.e., smaller $K$ from eq. (6)) such as the Entner–Doudoroff pathway [120]. By doing so they are able to consume exergy resources at a faster rate and outcompete other organisms. However, by utilizing these alternative catabolic pathways, they also dissipate more exergy as heat (i.e., they are less efficient) [121] and are thus less fit in exergy-limited environments. As a result of the selection for thermodynamic optimality, there is a change in the relative gene frequencies for power and yield pathways depending upon the amount of available chemical exergy in the surrounding environment [120]. Thus, empirically, it is observed that under conditions of surplus exergy, ecological assemblages will be selected for the ability to acquire and process that exergy to maximize power, whereas in environments with limiting exergetic resources, communities will form, which decrease dissipation, thereby maximizing yield.

5.3 Minimum dissipation and maximum power in evolutionary processes

Due to the tradeoffs of power and yield, theory predicts that systems should be selected for maximum power under conditions of surplus exergy and selected for minimum specific dissipation if exergy is limiting. Ken Spitze’s [122, 123, 124] work on the life history in Daphnia pulex presents several accounts of organisms being selected for these two attributes under the appropriate energetic regimes. Spitze [123] found that, if $D.\ pulex$ populations were exposed to predation pressure from Chaoborus larvae, they were younger and larger at the time of first reproduction and were also more fecund. This raised the following question: Why do these “superfleas” not outcompete the less fecund “unterfleas” and quickly rise to fixation in the population? Van
Table 2: Examples of the application of thermodynamics to biological systems at multiple scales.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In feeding experiments with mice, the animals had an increase in maximum</td>
<td>Hammond and Diamond, 1997 [128]</td>
</tr>
<tr>
<td>sustainable metabolic rate as well as basal metabolic rate when fed ad lib</td>
<td></td>
</tr>
<tr>
<td>rather than being fed controlled portions at specific times.</td>
<td></td>
</tr>
<tr>
<td>In snakes, active foraging strategies have a lower metabolic cost than</td>
<td>Secor and Diamond, 1998 [129]</td>
</tr>
<tr>
<td>sit-and-wait predators when meals are frequent and small, but a higher</td>
<td></td>
</tr>
<tr>
<td>metabolic cost when meals are infrequent and large.</td>
<td></td>
</tr>
<tr>
<td>Microbes in areas with higher chemical exergy in the form of bioavailable</td>
<td>Roach et al., 2017 [121]</td>
</tr>
<tr>
<td>carbon have a higher thermodynamic power output, as measured by isothermal</td>
<td></td>
</tr>
<tr>
<td>calorimetry.</td>
<td></td>
</tr>
<tr>
<td>If a resource (e.g., a limiting nutrient for a plant) increases in</td>
<td>Vollenweider, 1975 [130]</td>
</tr>
<tr>
<td>abundance, it will typically cycle faster.</td>
<td></td>
</tr>
<tr>
<td>For lakes with significantly different levels of eutrophication the</td>
<td>Salomonsen, 1992 [131]</td>
</tr>
<tr>
<td>exergy-to-power ratio is approximately the same in all cases.</td>
<td></td>
</tr>
</tbody>
</table>

Noordick and de Jong [125, 126] hypothesized that these results could occur if there was genetic variation for both power and yield. This would mean that superfleas are only super when resources are high, but when resource availability is low, they are less fit than the unterfleas. Tessier et al. [127] went on to show that the super phenotype was indeed only more fit when resources are high, as the smaller, less fecund phenotype was able to reduce available resources to a lower concentration than the larger, more fecund phenotypes. Here they [124, 127] demonstrate that the increased efficiency of the unterfleas allows them to have a fitness advantage when resources are limiting. For more examples of empirical studies and phenomenological support for the application of thermodynamics to biological processes, see Table 2.

6 Discussion

The thermodynamics of living systems as informed flows of exergy [132, 133, 134] gives a physical foundation for evolution by natural selection [24, 25, 89, 133, 135, 136, 137, 143]. How much of these flows is dissipated in any particular part of the process depends upon the information passed through time and the available exergy flow pathways. Any improvement in efficiency or an increase in speed of operation (power) without other costs will be favored in any biological entity and in any environment. However, FTT demonstrates that there is an explicit tradeoff between efficiency and rate that leads to a set of Pareto optima which optimize some combination of power and yield under the given conditions of the system. Which Pareto optimum is chosen depends upon the local orientors.

Thermodynamic functions provide important orientors of biological change at all scales, including ontogeny, development, succession, and evolution. The central notion of the orientor has its beginnings in the fields of systems analysis, cybernetics, and complexity science. The basic idea is that all dissipative systems are able to build up gradients and macroscopic structure from “disordered” elements in a system. Thus, similar characteristics can be observed in many different systems, and by utilizing these common aspects of system evolution, systems can be described as being oriented toward specific points in state space. The corresponding state variables that are used to elucidate these dynamics are termed orientors and their technical counterparts in in silico modeling are termed goal functions. See Appendix A (Table 3) for a list and description of biological orientors.

Which orientors are appropriate and best describe the trajectory of the system will depend upon the temporal and spatial scales of the process. In general, initial stages and short time scales are characterized by maximizing exergy acquisition and power in line with Lotka’s power law [24, 25] and Odum’s maximum emergy principle [138], whereas later stages are represented by efficiency serving to maximize yield and minimize dissipation. However, other constraints and objectives will also determine the exact Pareto optimum of power and yield. The theorems relating thermodynamic length to minimum entropy production (see eq. (6))
[93, 94, 139] demonstrate that dissipation can be decreased by adding additional stages [95] or by making the existing stages closer to equidistant [97, 140]. Understanding the ways in which biological systems add exergy extracting steps to processes and control the thermodynamic path length between steps aids in predicting which evolutionary and ecological strategies are most likely to be explored. Thus, FTT enables additional predictive power over current evolutionary and ecological theory.

This theoretical framework could prove particularly useful for pre-biological evolution and origin of life research. As the presiding evolutionary dynamics of variation, constraint, and selection are established during emergence of life, their thermodynamic basis can be seen most clearly by examining them apart from the specifics of biological organization [48, 135]. The principles that govern biological emergence carry through to the rest of evolution. Thus, understanding the emergence of life from pre-biotic chemical systems can bring to light new organizing principles of macroscopic biological change. For astrobiology the thermodynamic generalities of living systems should prove fruitful in the search for extra-terrestrial life [141]. Lastly, thermodynamic approaches provide insight into how trophic dynamics are directing evolutionary trajectory and ecological succession.

Future research should focus on merging the fields of developmental biology, ecology, and evolution with thermodynamics, information theory, cybernetics, and autopoiesis, to better understand the relation between boundary conditions, distance from equilibrium, dissipative rates, system complexity, exergy flow, and entropy production. The use of calorimetry, isotopic labeling, and manometry will aid in developing whole-system throughput matrices with data on primary production, respiration, biomass accumulation, dissipation, and turnover rates to provide more detailed quantitative mechanisms, models, and predictions for biological processes.

7 Conclusion

By utilizing non-equilibrium thermodynamic theory, biological phenomena at multiple time scales (Figure 1), ranging from development to succession and evolution, can be understood as physical processes (Figure 2). In order to apply thermodynamic theory to biological processes, biological systems should be treated as hierarchical networks of informed exergy flows. These exergy flows are informed by the passage of information from a previous time. It is this passage of information through time that allows living systems to build up technology that extracts work from exergy streams. This passage of information through time is also what allows for evolution by natural selection. Thus, from a thermodynamic perspective, natural selection acts as a control, which continuously optimizes the tradeoff between power and yield (i.e., a Pareto optimum of power and yield) (Figure 3). It is emphasized that biological systems are not driven by the production of entropy, but rather by the flow of exergy through the informed pathways of the system. The entropy production is the generalized friction, which is minimized subject to the constraints of finite time and resources.

The application of FTT to living systems demonstrates that there is an explicit tradeoff between power and yield in biological processes. Utilizing theorems from FTT concerning thermodynamic length, process rate, and dissipation (see eq. (6)), we develop hypotheses concerning the effects of this tradeoff on ecological and evolutionary systems and offer a number of relevant studies, across several biological scales, as supporting examples (Section 5 and Table 2).

To quantify the tradeoff between power and yield, we have introduced the notion of compensated heat and waste heat (Box 3). Compensated heat is the dissipated heat necessary to achieve a desired rate or meet a constraint. Waste heat is any exergy that is dissipated as heat which does not provide a compensatory benefit (Figure 4). Within this framework, there are characteristic generalities of non-equilibrium thermodynamic systems across a variety of timescales. We emphasize that biological systems do not maximize power or minimize dissipation, but rather optimize the tradeoffs between power and yield within the functional control region of operation. However, in general, living systems tend to maximize power over short time scales and when resources are high, and tend to minimize dissipation over longer timescales or when resources are low. This can be seen in the strategies employed at several biological scales, including organismal development,
ecological succession, and long-term evolution [142]. Better understanding the roles of compensated heat and waste heat in biology will provide novel insight into the underlying physical mechanisms driving biological processes at multiple temporal and spatial scales.

**Acknowledgment:** The authors would like to acknowledge the San Diego State Biomath Group and the Energy Landscapes Workshop at the Telluride Science Research Center for their insightful discussions and valuable feedback concerning this work.

**Funding:** The authors thank the National Science Foundation for providing funding which supported this research in the form of a Graduate Research Fellowship provided to T.N.F.R.

---

**Appendix A. Table of biological orientors**

**Table 3: Orientors for evolutionary and ecological trajectories.** Factors in italics are components of the orientor, which may serve as proxies to measure the orientor. Abbreviations are as follows. TST: Total System Throughput, TSE: Total System Export, TSS: Total System Storage, AMI: Average Mutual Information.

<table>
<thead>
<tr>
<th>Orientor</th>
<th>Description</th>
<th>Relevant literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum power</td>
<td>Systems develop to maximize the flow of energy. In network context this amounts to maximizing TST, that is, maximizing the total energetic flow into or out of all compartments of a system.</td>
<td>Lotka, 1922a,b; Lotka, 1925; Odum and Pinkerton, 1955</td>
</tr>
<tr>
<td>- <strong>TST</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- <strong>Turnover rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- <strong>Biomass</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- <strong>Production: biomass</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(minimizes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- <strong>Production: respiration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(approaches unity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum empower (emergy)</td>
<td>Systems organize in a manner that maximizes emergy flow and storage.</td>
<td>Odum, 1988; Brown, Odum, and Jorgensen 2004</td>
</tr>
<tr>
<td>- <strong>TSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy production</td>
<td>Systems will organize to maximize entropy production early in development and then switch towards minimizing entropy production at later stages of development.</td>
<td>Aoki, 1989</td>
</tr>
<tr>
<td>- <strong>TSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum ascendency</td>
<td>Systems tend to maximize ascendency (i.e., the total system throughput times the AMI) over time.</td>
<td>Ulanowickz, 1980, 1986, 1997, 2006; Patricio et al., 2004</td>
</tr>
<tr>
<td>- <strong>TST</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- <strong>AMI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum storage</td>
<td>If multiple paths are available to move away from thermodynamic equilibrium the one offering the greatest amount of stored exergy will be selected.</td>
<td>Jorgenssen 1992, 1997, 2011</td>
</tr>
<tr>
<td>Minimum empower: exergy ratio</td>
<td>Systems develop in a way that yields the minimum cost (TST) to produce a given amount of organization (TSS).</td>
<td>Bastianoni and Marchetini, 1997; Bastianoni, 1998</td>
</tr>
<tr>
<td>- <strong>TST</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- <strong>TSS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum dissipation</td>
<td>Systems develop in a manner that will eventually maximize the dissipation of energy gradients.</td>
<td>Schneider and Kay, 1990, 1994a,b, 1995, 1996; Prigogine, 1955; Prigogine and Stengers, 1984; Brooks and Wiley, 1986</td>
</tr>
<tr>
<td>- <strong>TSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum cycling</td>
<td>Energy flow in systems will generate cycling which is a form of organization. As systems grow and develop they will tend to maximize their energy flowthrough and concomitantly their cycling.</td>
<td>Morowitz, 1968</td>
</tr>
<tr>
<td>Maximum residence time</td>
<td>Residence time is a measure of the fraction of the TST that remains as storage. As systems develop they will organize to maximize the residence time of stored exergy.</td>
<td>Cheslak and Lamarra, 1981</td>
</tr>
</tbody>
</table>
Table 3: (continued)

<table>
<thead>
<tr>
<th>Orientor</th>
<th>Description</th>
<th>Relevant literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum specific dissipation</td>
<td>As systems approach steady states they will organize to minimize the dissipation per unit mass.</td>
<td>Prigogine, 1947, 1955; Prigogine and Wiame, 1946 Christensen, 1994</td>
</tr>
<tr>
<td>Relative ascendency</td>
<td>Relative ascendency is the ascendency of a system relative to its maximum ascendency (developmental capacity). As systems mature relative ascendency tends to approach 1.</td>
<td>Christensen, 1994</td>
</tr>
<tr>
<td>Decreased overhead</td>
<td>As ecosystems near maturity the overhead tends to decrease toward a minimum.</td>
<td>Rutledge et al., 1976</td>
</tr>
</tbody>
</table>

References


[34] Henri Bénard, Les tourbillons cellulaires dans une nappe liquide propageant de la chaleur par convection: en régime permanent, Gauthier-Villars, 1901.


[60] I. Prigogine 1980 From being to becoming: time and complexity in the physical sciences. W. H.


