Thermodynamics in finite time

Asking how well systems can perform if they are to deliver power, not just energy, leads to investigations both in abstract, fundamental thermodynamics and in almost-applicable physics, such as determining the optimal motion of a piston.

Bjarne Andresen, Peter Salamon and R. Stephen Berry

Until the 19th century, technology was essentially the domain of skilled artisans and constructors who relied on practical experience to design and build their machines. One of the first efforts to use physical theory to study the functioning of machines was undertaken by the French engineer Sadi Carnot. Motivated by the concern of the French about the superiority of British steam engines, he undertook a systematic study of the physical processes governing steam engines, resulting in his remarkable paper *Reflexions sur la puissance motrice du feu* (On the Motive Power of Heat) published in 1826. Among the earliest successes of this new science, thermodynamics, was the formulation of criteria describing how well real processes perform in comparison with an ideal model. Carnot showed that any engine, using heat from a hot reservoir at temperature T_h to do work, has to transfer some heat to a reservoir at lower temperature T_l, and that no engine could convert into work more of the heat taken in at T_h than the fraction

$$\eta_C = 1 - (T_l / T_h)$$

known as the Carnot efficiency.

Other criteria for the performance of heat engines emerged with the introduction of the concepts of energy conservation and of thermodynamic potentials: Hermann von Helmholtz's definition of a free energy H, Josiah W. Gibbs's concept of "available work" or availability A, the Gibbs free energy G, the effectiveness ϵ (the ratio of the actual work supplied by a work source to the change in availability of the system), and others. All these criteria have long been common currency for thermodynamic studies in physics, chemistry, and engineering. They all share one characteristic: The ideal to which any real process is compared is a reversible process.

Our research in thermodynamics was stimulated by the challenge to reduce the use of energy in heat engines. In this context, we can question the usefulness of comparing real processes to reversible processes—which, after all, take an infinite time to complete. For example, are we interested in a factory that makes automobiles infinitely slowly? The basic question is: Are reversible limits close enough to real performances to be useful in guiding the improvement of processes? This question leads almost immediately to the consideration whether it is possible to find more realistic limits to the performance of real processes. Can we find bounds for how well those processes can operate in finite time? Can we use such bounds to find better criteria of merit useful in evaluating real processes? In general terms, finite-time thermodynamics is concerned with how constraints on time or rate affect performance. (See figure 1.)

The question of how well a system can perform in finite time has led to a wide spectrum of new scientific inquiries, ranging from existence theorems and basic problems, such as finding limits on entropy production and defining adequate models for real systems, to some challenging questions in almost-applicable physics, for example determining the optimal time path of the piston in an idealized automobile engine, or designing new kinds of engines based on dissipative processes. Let us briefly consider the latter two problems as examples of practical applications of our approach.

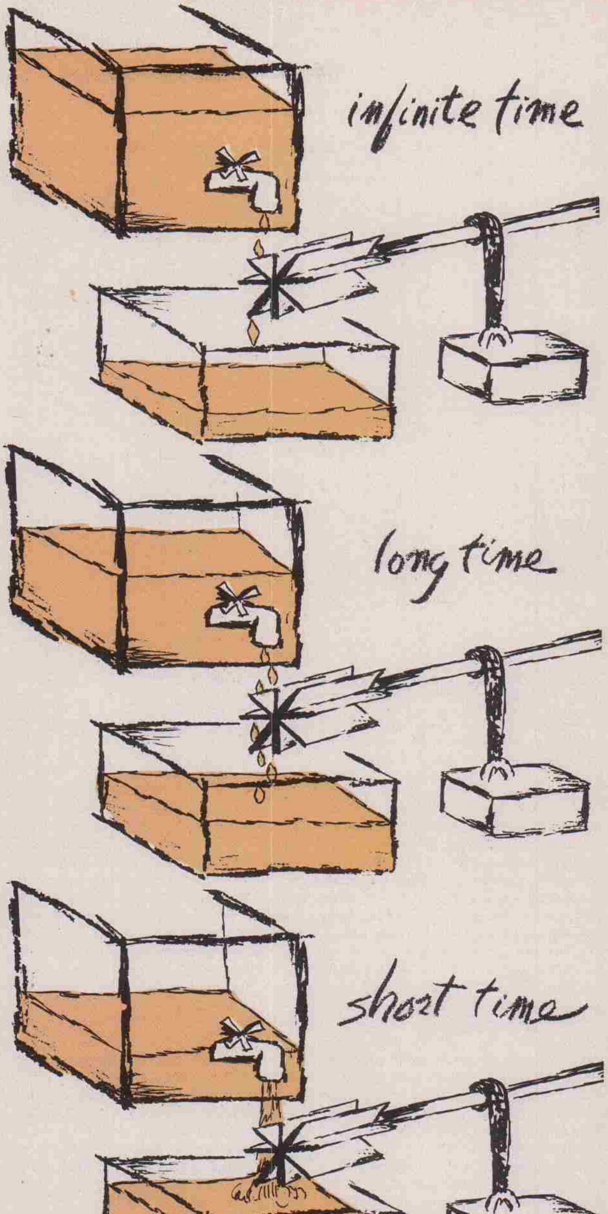
Bjarne Andresen is at Physics Laboratory II of the University of Copenhagen, Denmark; Peter Salamon is in the mathematical sciences department at San Diego State University; R. Stephen Berry is in the department of chemistry at the University of Chicago.

Engine design

As a model for the Otto cycle[1] we chose an engine with friction and a finite heat leak proportional to the exposed cylinder area. The engine operates at constant period and constant fuel consumption per cycle. We optimized the time path of the piston to yield the maximum work per cycle, keeping constant engine parameters such as the friction constant, the heat leak coefficient, and the period. We considered a four-stroke cycle of intake, compression, power, and exhaust strokes. The intake, compression, and exhaust strokes have minimum losses when they are run at constant velocity for as long as possible, preceded and followed by acceleration (or deceleration) at the maximum possible rate to reach the optimum constant velocity.

The most interesting part of the cycle is the power stroke. In real automobile engines, heat leakage, especially during the hottest portion of the power stroke, is the most important source of inefficiency. Consequently, in the optimum path, the piston has to accelerate as fast as possible to a high expansion velocity during the power stroke. If the acceleration could be infinite, the optimum velocity at the beginning of the power stroke would be about twice the maximum velocity of the piston in a corresponding conventional engine. The velocity of the optimized power stroke then decreases with time (or piston position), quickly becoming linear with piston displacement. Figure 2 shows the velocity and position of the piston for an optimized and a conventional cycle. The optimized cycle has a maximum acceleration of 5×10^4 m/sec^2, about ten times that of the piston in a corresponding conventional en-

 0031-9228 / 84 / 0900 62-09 / $01.00

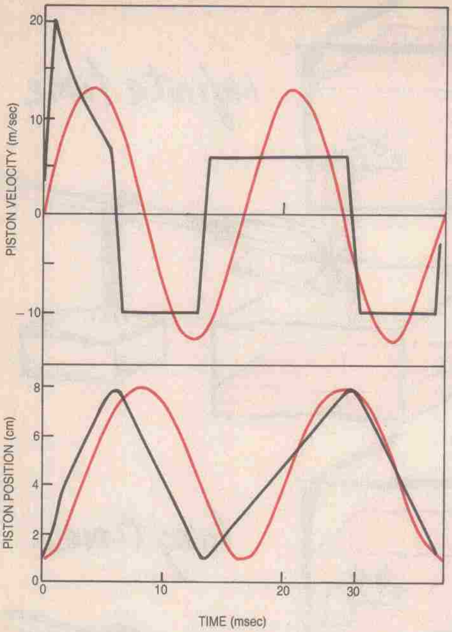


The cost of speed. A weight is lifted by letting falling water turn a paddle wheel. With an infinite amount of time available, the machine can operate reversibly with no losses and maximum efficiency: All the available energy can be converted into work. However, if there is only a finite time available, losses are unavoidable and the machine operates irreversibly. As the operation becomes faster the losses increase; at some finite speed the machine operates at maximum power, but with an efficiency less than the thermodynamic maximum. Figure 1

gine.

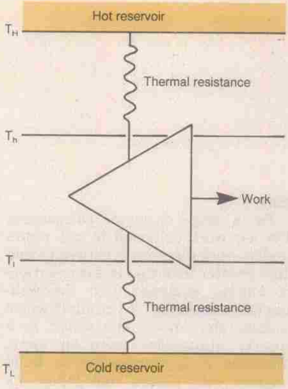
For a range of engine parameters, the net work delivered to the piston and its work reservoir is between 8 and 15% greater than that in a conventional engine, equipped with the well-known connecting rod-crank-flywheel linkage that moves the piston in a nearly sinusoidal path in time. Whether it is worthwhile to build engines with pistons that follow approximately optimized trajectories is still an open question.

Our second example is a new kind of engine based on dissipative processes. Recently John Wheatley and his coworkers[2] have conducted a set of experiments on an intrinsically irreversible acoustic heat engine that transforms a temperature gradient along a tube into acoustical energy, or vice versa. This engine exemplifies the class of systems that do not have a reversible limit in that they deliver zero work per cycle in that limit. In its most basic form, the heat pump consists of a thin tube, closed at one end, and containing a piston in the other. When the piston is moved back and forth at a suitable rate, the gas heats adiabatically during compression, while being moved toward the closed end of the tube, delivering heat to the walls of the tube. During expansion the gas cools, but is moved farther away from the closed end of the tube. The net effect is the transport of heat from the open end to the closed end of the tube. Thermal equilibration of the gas with the walls is of course irreversible. However, if one tries to remove this irreversibility by operating the device slowly, the temperature difference between the gas and the wall is reduced and the heat transfer reaches zero because the gas will equilibrate with the wall during compression. Such systems are intrinsically irreversible:



Optimized Otto cycle. The piston in an internal combustion moves in a sinusoidal path, as determined by the crankshaft (color). However, one could increase the engine's efficiency up to 15% by extracting work rapidly before the hot combustion gases lose heat to the cylinder walls and by keeping the piston speed as constant as possible to reduce frictional losses (black curve). Figure 2

Endoreversible engine. In this simple model of an irreversible engine all the losses are associated with the transfer of heat to and from the engine; there are no internal losses within the engine itself. Because of the finite conductivity of the heat-transfer material the engine operates not between T_H and T_L but between T_h and T_l; the temperatures T_h and T_l depend on the rate of heat flow, and thus on the power output of the machine. The efficiency of the engine thus depends on its power output. Figure 3



They do not have reversible counterparts. In the acoustic engine the irreversibility itself provides the necessary phase difference between the gas motion and temperature change. An electronic analog is the RC oscillator where a dissipative resistor replaces the lossless self-inductance in an LC oscillator.

In the remainder of this article, we emphasize the general ideas and physical principles of thermodynamics with constraints on finite-time operation; we conclude with a brief glimpse of a few more applications of the approach: determination of the maximum power obtainable from endoreversible engines (that is, engines for which the only irreversibilities are located in the coupling of the engine to its surroundings), from engines with time-dependent energy sources; and from engines using finite heat reservoirs. Other applications are the optimization of the time paths to maximize the mean power delivered by an Otto cycle engine (described above) and by an external combustion engine; the maximization of efficiency and of power in a system in which an exothermic chemical reaction supplies energy to drive a second reaction; and finally, the theory of an engine driven by light absorption, whose irreversibilities are, like those of the thermoacoustic engine described above, inherent to the operation of the engine. For each system, work per cycle becomes zero for infinite cycle periods, in contrast to the systems that are capable of producing work even when they are operated reversibly.[3]

Formulation of the theory

We can clarify the role of time and rate constraints in thermodynamics by asking a sequence of questions:

▶ What are the necessary and sufficient *conditions for the existence* of quantities—extensions of conventional thermodynamic potentials—whose changes give the extremal values of the process quantities of heat or work that may be exchanged during a process, when the constraints on that process include constraints on the rate or duration of the process? Naturally, these generalized thermodynamic potentials should be functions of conventional thermodynamic variables of state. They necessarily depend on the constitutive parameters and the rate or duration, but they also depend on one or more time-response parameters of the system. After all, if we are going to squeeze more physics out of the problem, we have to expect to put more physics into it.

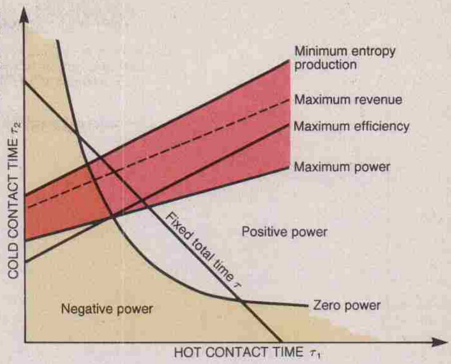
▶ Given the conditions for the existence of generalized thermodynamic potentials for finite-time processes, are there *algorithms for evaluating* them or their changes? When we do evaluate them, what do we learn about the price we pay for operating processes faster or slower?

▶ Can we find the idealized but finite-time *pathways* that will yield the extremal work or heat exchange given by the appropriate generalized potential, and how can we use these results to improve real processes?

▶ What are the operational differences among processes of the same kind but running so as to *optimize* different quantities?

It is useful to point out a major difference in viewpoint between work understood as finite-time thermody-

Optimizing an endoreversible engine. The engine of figure 3 spends time τ_1 in contact with T_H and τ_2 in contact with T_L; the time taken to perform work and go through the remainder of its cycle is immaterial, as those operations are reversible. The best choice of parameters depends on what one choses to maximize. Depending on the prices involved, the line for maximizing revenue can fall anywhere within the colored band. Figure 4



namics and the established body of research called "irreversible thermodynamics." Both deal with the extension of traditional equilibrium thermodynamics to systems with irreversibilities, and both are concerned with how those irreversibilities affect the behavior of processes. However, irreversible thermodynamics addresses a question different from that of finite-time thermodynamics: What are the equations of motion of the thermodynamic state variables of a system, and how can we solve them? Finite-time thermodynamics begins with the four questions formulated above, which may be summarized as: How do constraints on time or rate affect the net process variables—and therefore the performance—of a process? The focus of irreversible thermodynamics on equations of motion leads naturally to a formulation in terms of differential equations and to examination of the local, differential behavior of systems. The focus of finite-time thermodynamics on the net changes of process variables leads to variational principles and global descriptions of systems. There are of course many points of contact between the two approaches, but as yet only the simplest, most superficial relationships between the two have been explored. As some of the later examples will show, the viewpoint of finite-time thermodynamics sometimes makes it easier to address questions generally. For example, the global view may help to expose the extent to which assumptions of linear behavior limit the applicability of an analysis; the global view may even lead to more general approaches.[4]

The approach of finite-time thermodynamics also differs in two ways from that of the practicing engineer. The engineer optimizes a model that is as detailed a representation as possible—typically an elaborate and specific simulation—of the particular system he wants to build or use. Our approach, by contrast, aims to isolate how the dominant temporal characteristics of each broad class of processes set limits on the performance of that class. Secondly, the engineer typically optimizes the values of key parameters of the apparatus, whereas finite-time thermodynamic optimizations, at their most elaborate, find the temporal path that maximizes the performance of a process.

The construction of generic models to represent broad classes of processes is central to finite-time thermodynamics. Each generic model should contain all the important qualities of the type of real system studied, but not the individual details which would obscure the physical content and make calculations very difficult or impossible. We are already used to such models in traditional thermodynamics: the Carnot engine is, for example, the highly idealized reversible representation of all heat engines. Finite-time thermodynamics retains the same philosophy of model construction, but makes the models somewhat more realistic. The first "improvement" of the Carnot engine is the endoreversible engine shown schematically in Figure 3.

Generalized potentials

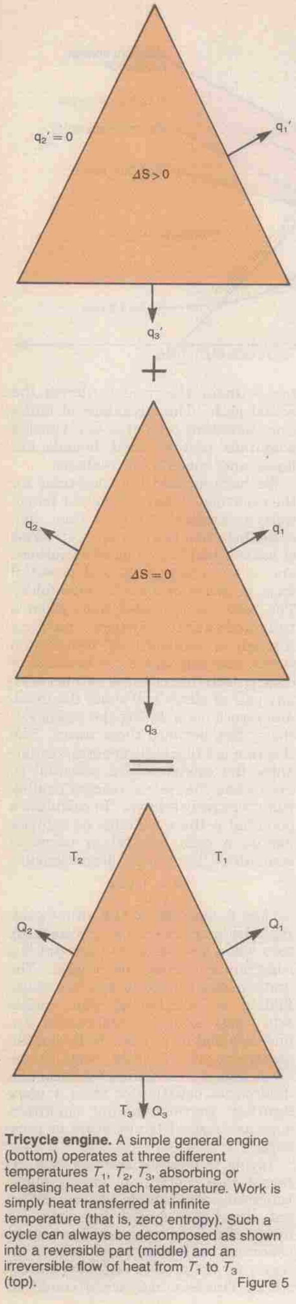
A potential is a measure of the capability for a system to do work; a thermodynamic potential is a generalized potential in the sense that heat, as well as any mechanical, electromagnetic or other reversible source, may be used as a source of the work. The traditional thermodynamic potentials are state functions defined in such a way that the difference in the potential between any state A and any other state B of the system considered is the maximum (and therefore reversible) work that the system can produce during any process that carries it from state A to state B under given constraints. For example, the internal energy U applies to adiabatic processes, the Helmholtz free energy H applies to isothermal–isochoric processes while the Gibbs free energy G applies to isothermal–isobaric processes. All of these potentials may be generalized,[5] under suitable conditions, to describe the capability of a system to do work under arbitrary constraints, for example, within a specified interval of time. The utility of these generalized potentials thus is the same as of those we have become accustomed to in traditional thermodynamics, namely, they provide upper bounds to work production without the need to derive the actual path. The advantage of finite-time potentials compared to reversible potentials is that their bounds are lower and hence more realistic.

We have derived the conditions for the existence of thermodynamic potentials generalized to finite-time processes both in a familiar form, in terms of conventional calculus of variations, and in a more general and powerful form in terms of abstract manifolds.[5] The basic result states that, given a thermodynamic system passing through a sequence of equilibrium states, one can identify a function of state $\wp$ such that changes in $\wp$ between any pair of states will equal the maximum work done during the passage of the system between those states. The theorem is sufficiently general to guarantee the existence of a potential $\wp$, even when the system reaches equilibrium relatively seldom. To calculate a potential $\wp$ the work must be expressible as a path integral in terms of generalized forces and displacements

$$W = \int \mathbf{f} \cdot d\mathbf{x}$$

To find $\wp$, one adds to $\mathbf{f} \cdot d\mathbf{x}$ an integrating term $\mathbf{g} \cdot d\mathbf{y}$, where $d\mathbf{y}$ is necessarily zero when one takes into account the constraints defining the process. The mathematical problem has two steps: finding a function $\mathbf{g}$ that makes $\mathbf{f} \cdot d\mathbf{x} + \mathbf{g} \cdot d\mathbf{y}$ an exact differential $d\wp$, and then finding $\wp$ itself. Note that the constraint $d\mathbf{y} = 0$ could itself be derived from a condition in the form of a differential equation or from a more familiar thermodynamic condition such as constant temperature or pressure.

Trying to describe systems with internal rate constraints among quantum states, for example, between different vibrational modes, has made us aware of the importance of the second condition—of having a well-defined integral for the process variable (usually work). This seemingly trivial condition



Tricycle engine. A simple general engine (bottom) operates at three different temperatures T_1, T_2, T_3, absorbing or releasing heat at each temperature. Work is simply heat transferred at infinite temperature (that is, zero entropy). Such a cycle can always be decomposed as shown into a reversible part (middle) and an irreversible flow of heat from T_1 to T_3 (top). Figure 5

becomes more important when one tries to describe the extraction of work from a system at a rate faster than that system can equilibrate internally. In such a case, thermodynamic variables lose their meaning, and one must go to a definition of work in terms of energy transfer at the microscopic level. This approach can sometimes be too complex to be useful. Nevertheless, work and availability can be defined for some systems, such as simple lasers whose operation depends on changes in populations of specific quantum states.[6] Another approach relies on the "maximum entropy" formulation used in information theory for defining thermodynamic variables: If P_j is the probable fractional population of the jth level, then one postulates that the distribution over all the quantum states of the system takes on that form which maximizes the quantity $-k\Sigma_j P_j \ln P_j$, subject to whatever constraints are given. This maximized sum is then identified as the entropy S. Friedrich Schlögl and later Raphael Levine established the relation between the maximum-entropy approach and thermodynamic availability[7] as part of a broad effort to develop this method.

The optimal path

In some cases it may not be enough to know what is the maximum work that can be extracted during a given process—calculated, for example, from the generalized potentials described above. One may also want to know how to extract this maximum work, that is, to specify the time path of the thermodynamic variables of the system. The tool for obtaining this path is the discipline called optimal-control theory.

To set up an optimal-control problem we must specify:

▶ The controls, that is, the variables that can be manipulated by the operator (for example, a volume, rate, voltage, heat conductance, and so on)

▶ The limits, if any, on the controls and on the state variables (for example, to avoid unphysical situations, such as negative temperatures and infinite speeds)

▶ The equations that govern the time evolution of the system (usually differential equations describing heat-transfer rates, chemical-reaction rates, friction and other loss mechanisms)

▶ The constraints that are imposed on the system: what is conserved, what is held constant, what processes are reversible, and so forth (expressed as differential or algebraic relations, representing constraints acting at every instant, or as integral or global relations, representing constraints acting over the entire interval)

▶ The desired quantity to be maximized, called the objective function and usually expressed as an integral

▶ Whether the duration of the process is fixed or part of the optimization.

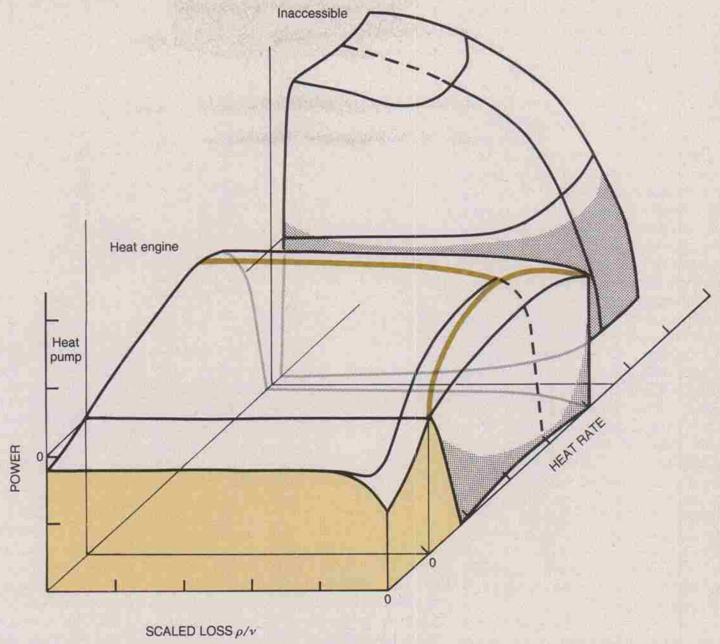
Maximizing the objective function with the usual methods of the calculus of variations leads to a set of coupled, nonlinear differential equations. These are usually so complex that a qualitative analysis and a numerical solution are the only hope. Thus, answering the more exacting question about the optimal time path rather than the standard question about maximum performance requires a considerably larger computational effort. On the other hand, once the time path is calculated, all other thermodynamic quantities may be derived from it, much like the wave function is the basis of all information in quantum mechanics. Many results we describe in subsequent sections of this article have been obtained by optimal control theory.

Criteria of performance

The earliest criterion of performance for engines, the efficiency, was used to measure how much water could be pumped out of a mine by burning a ton of coal. It has subsequently been generalized to provide a criterion of performance for all heat engines: work produced per heat energy input. Other criteria, such as effectiveness, can be defined that involve loss in availability or changes in other thermodynamic potentials.

While the traditional criteria of performance generally consider the ideal process as a reversible one, some investigators have considered processes operating at nonzero rates. J. Geusic, E. O. Schulz-DuBois, and H. E. D. Scovil[6] treated the three-level laser with a thermodynamic formalism that looks amazingly similar to one we later devised independently[8] for macroscopic processes with heat leaks, friction, and finite rates of heat transfer. The laser analysis maximized the work that could be produced as light energy when the ratios of level populations are used to define the reservoir temperatures.

Another treatment[9] has evolved into almost a classic paradigm of systems operating in finite time: Frank L. Curzon and Boye Ahlborn considered a Carnot engine with the ingenious variation that it is linked to its reservoirs through finite heat conductances (figure 3). Curzon and Ahlborn asked, what process maximizes the power produced by the engine. Any reversible system produces zero power, of course, and maximizing power forces us to deal with systems operating at finite rates. Operating reversibly, the efficiency of this engine is just the Carnot value of $1-(T_l/T_h)$; but Curzon and Ahlborn were able to show that when the system operates to produce maxi-



Power output of an engine as a function of the heat rate and its internal friction. The heat flows to and from the engine through materials of finite conductivity (as in figure 3), and there are frictional losses within the engine, taken as linear in the speed of the engine. At negative heat rates, the engine functions as a heat pump. At positive rates it delivers power, up to a maximum; heat rates above the maximum are physically inaccessible. When the friction ν is small compared to the thermal resistance ρ (to the left) there is a single—and in fact constant—heat rate that maximizes power. When the friction is large there are two possible heat rates (and efficiences) that yield the same power. Figure 6

mum power the efficiency of the engine is only $1-\sqrt{(T_1/T_h)}$. In such an engine a maximum for the power can only be achieved by paying for it in lowered efficiency. Curzon and Ahlborn also pointed out that typical power plants are properly so named: They operate much nearer the point of maximum power than the point of maximum efficiency.

Other criteria of performance are the rate of entropy production and the rate of loss of availability. The concept entropy production was introduced in the earliest thinking about irreversible thermodynamics—but more from the differential, local and instantaneous viewpoint than from the global, integral view of entire optimized processes. Under some circumstances, optimizing one of these quantities is equivalent to optimizing another. For example, in those cases in which the irreversibilities can be represented as spontaneous heat flows, minimizing the entropy production turns out to be equivalent to minimizing the loss of availability.[4] One can sometimes say a good deal about the possible behavior of a system even if one knows only that it operates to optimize some (unknown) function from a known class. While such an approach has obvious, tantalizing possibilities for biological systems, the only case in which it has been applied is in a description[10] of maximizing the revenue from a power-generating plant. All solutions to the maximized-revenue problem for an endoreversible engine with finite heat conductances are bounded on one side by the solutions to the maximum-power problem and on the other side by the solutions corresponding to minimum loss of availability as illustrated in figure 4. The optimal controls of both extremes have been worked out in detail for endoreversible heat engines.[4,11]

The tricycle formalism

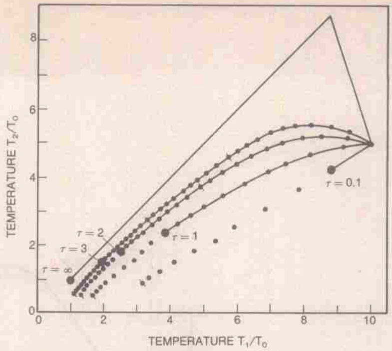
An entirely different way to assess the cost of finite-time operation is the "tricycle" formalism,[8] a construction based on conservation equations for the process in question. We represent a heat exchange-system schematically, as in figure 5, by a triangle with heat flows Q_1, Q_2 and Q_3 into reservoirs with temperatures T_1, T_2 and T_3. A conventional heat engine or refrigerator is a special case of this scheme, with one of the temperatures, T_1 say, infinite, so that zero entropy flow is associated with the energy flow Q_1, which can thus be identified as work. In general, one can divide a tricycle process into its reversible part with zero entropy production, $\Delta S = 0$, and a totally irreversible component, as indicated in figure 5. One learns nothing new, of course, from such a decomposition *per se*, but by including specific loss mechanisms, such as heat resistance, friction and heat leaks, one can deduce the rate dependences of these irreversibilities. It may seem that these three loss mechanisms should be treated individually, but they are in fact interdependent and can be determined simultaneously.

A convenient form to present the results is a contour diagram or a three-dimensional plot, such as figure 6, where the height represents power output for a heat engine or heat pump. (We have plotted reduced variables: the ordinate is a measure of heat rate, and the abscissa of friction, both measured relative to thermal resistance.) Observe that the heat-rate axis is divided into three distinct regions: The system operates as a heat pump for all negative values and as a heat engine for positive rates, but only up to a certain limit; operation at larger values is impossible because such large heat rates would require negative temperatures to overcome the resistances.

The loci of maximum power are indicated in color, and for negligible friction (to the far left) there is indeed a single maximum power the engine can deliver, as we discussed earlier. As friction increases, a classical bifurcation splits the locus of maximum power into two, with a minimum in between. In this region there are two quite different heat rates, each of which produces the same maximum power—but obviously at vastly different thermal efficiencies. The bifurcation divides the heat-engine region into distinct regions: a low-friction region dominated by heat resistance, and a high-friction region. Because there is no limit to how much power one can supply to speed up the pump, the heat-pump region (negative heat rate) shows no such structure.

The advantage of such diagrams over a single optimum-performance number is that one can also read the cost for off-optimum operation. For example, in the low-friction region of figure 6 it is very slight. One can, of course draw similar diagrams for the efficiency or any other objective function.

The tricycle formalism, with the reversible and irreversible contributions separated, allows one to calculate which are the most serious losses from irreversible processes. Certain processes, such as ordinary distillation, have unavoidable irrebersibilities built into them. By separating the tricycle for the unavoidable losses from that for excess losses, one can determine how much room there is for improvement: The cycle of excess losses can show how to improve the process itself; the cycle of unavoidable losses can show what improvements can result from a change in process.



Effect of time scale on engine operation. Two endoreversible engines extract heat from two finite reservoirs at T_1 and T_2; both engines reject heat to an infinite low-temperature reservoir at T_0. The hot reservoirs are also coupled to each other via a heat-conducting medium. The object is to maximize the amount of work extracted during a time τ while the two reservoirs are also coming to equilibrium. The spots indicate the temperatures of the reservoirs at each 0.1 unit of time. Note that the evolution of the system depends on the value of the time τ. Figure 7

Finite-time availability

Two of the most recent, and the most powerful results obtained so far, relate to the available free energy in finite-time processes. One is a definition[12] of "finite-time availability." Gibbs originally defined the availability A of a system in contact with given surroundings to be a state function such that the decrease in availability in going from state i to state f is the maximum (and hence reversible) work that can be extracted during this process. The finite-time availability $\mathscr{A}$ retains this property of telling us how much work a system can supply. When we add the restriction that the process must operate only during the interval $\tau = t_f - t_i$, then we define

$$-\Delta\mathscr{A} = W_{\max}(\tau) = \max\left\{ A(t_i) - A(t_f) - T_0 \int_{t_i}^{t_f} \dot{S}_{\text{tot}}\,\mathrm{d}t \right\}$$

where $\dot{S}_{\text{tot}}$ is the total rate of entropy production. In this expression the difference in traditional availability $A(t_i) - A(t_f)$ is of course the reversible work produced in a process from the initial to the final state; the last term represents the work lost to the environment by production of entropy (that is, the losses in the process). The search for a maximum be constrained to reach exactly a given final state at t_f (the initial state is always considered known), in which case ΔA is fixed and the optimization becomes one of minimizing the entropy production; the search may also include the final state in the optimization, in which case $\Delta\mathscr{A}$ must be evaluated by optimal control. If the final state is specified, a solution may not exist if τ is too short, because only a certain set of states can be reached from a given initial state in time τ.

The finite-time availability does not necessarily have ΔA as its limit for very long times, because the system may contain internal relaxation processes that remain irreversible even for very slow operation. If there is a direct heat leak from the system reservoir to the surroundings, as in the tricycle example above, then a long process time τ may even reduce $-\Delta\mathscr{A}$ to zero.

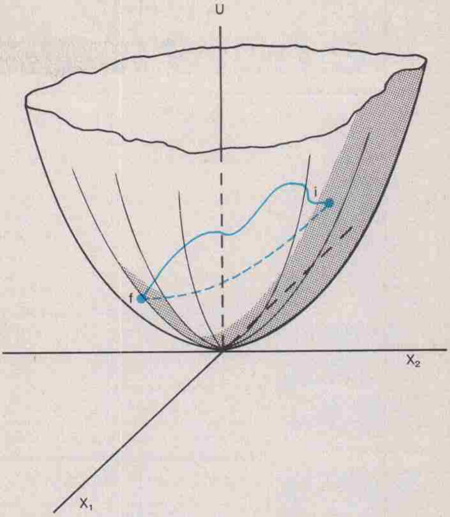
The finite-time availability is as general a quantity as the traditional availability, that is, it can be applied to any thermodynamic process. Figure 7 shows some results from a study[12] of a work-producing system with competing internal relaxation. The description used in this study is general enough to be applied, for example, to heat engines, internal molecular degrees of freedom, hydraulic systems, or chemical reactions, simply by suitably identifying the variables.

The other recent finding concerning availability is a very simple and general bound for the Gibbs availability A that must be lost in a process if a system is driven in a finite time τ via states of local thermodynamic equilibrium from an initial equilibrium state i to a final equilibrium state f. This bound[13] makes explicit the cost in availability of finite-time operation; it arises as a result of minimizing the "distance" traversed in going from i to f in the abstract space of the independent extensive variables of the system. Frank Weinhold[14] identified the matrix second derivatives of the internal energy U with respect to the extensive variables as a metric tensor on this space, that is, the distance between "points" separated by $\mathrm{d}X_i$ is given by

$$\mathrm{d}s^2 = \sum_{ij} (\partial^2 U / \partial X_i \partial X_j)\,\mathrm{d}X_i\,\mathrm{d}X_j$$

The newly derived bound is a function

Internal energy U of equilibrium states of a thermodynamic system as a function of extensive variables $X_1, X_2, \ldots$ forms a surface in a multidimensional space. One can define a metric on this surface such that the length of the path taken between two states of the system, i and f, is related to the loss of availability engendered by the process. Figure 8



of this distance: If the process takes place in time τ, then the dissipated availability $-\Delta A$ is bounded by the square of the length of the shortest path from i to f multiplied by ϵ/τ, where ϵ is a mean relaxation time of the system. If the process is endoreversible, the bound can be strengthened to $-\Delta A \geqslant L^2\epsilon/\tau$, where L is the length of the path actually traversed from i to f. The geometry is illustrated in figure 8.

Some specific systems

We now turn to some specific optimizations. Clearly we cannot list all the important contributions to this field—these include, for example, work by Michael Mozurkewich, John Ross, Robert Ross, Luigi Sertorio and Schlögl in addition to what we describe below.

Morton Rubin[15] applied the variational tools of optimal control theory to find the time path for the endoreversible engine of figure 3 that maximizes the power output. Without any restrictions as to the type of branches, limiting only the reservoir temperatures, heat conductances, and rate of change of volume, he found the optimal cycle to consist of six branches: two isotherms and four maximum-power branches, but no adiabats. The adiabats of a Carnot cycle are replaced by the maximum-power branches along which the volume changes at its maximum permissible rate; the working fluid is still in contact with one or the other heat reservoir. Thus the optimal cycle is never isolated from both its reservoirs. If the limitation on the rate of volume change is lifted, these new branches proceed instantaneously and thus become adiabats. In effect, the system operates as nearly like a Carnot cycle as its time constraints allow.

In a subsequent study[16] Rubin made his system more realistic by restricting the compression ratio, that is, the ratio between the largest and smallest volume, to a fixed value. With this constraint the number of branches in the optimal cycle increased to eight: two fixed-volume branches enter where each pair of maximum-power branches meet. These branches allow the working fluid to heat up or cool at its limiting volumes.

In most real heat engines the heat sources and heat sinks are not infinite. For example, the combustion products in gasoline and diesel engines have finite heat capacities and the exhaust gases should in principle be cooled completely to the surrounding temperature during the power stroke. In some engines magnetohydrodynamic pre-cycles and steam-turbine bottoming cycles are added to conventional heat engines to increase the heat energy used. However, maximizing the use of heat in such systems will probably require new cycles with heat-accepting (and possibly discharging) branches designed to match the heat capacity of the exhaust gas. Finite-time thermodynamics can provide such new cycles.[17]

Yehuda Band, Oded Kafri and Salamon[18] have designed a model that maximizes the power from an engine driven by a time-dependent source supplying heat at a rate $f(t)$. This thermal energy simulates a variable supply of energy to the working fluid either from internal degrees of freedom, as one finds in chemical reactions, or from an external source, as in laser-driven engines or external-combustion engines. The model has the working fluid coupled to a heat bath at temperature $T_{res}(t)$ through a coupling obeying Fourier's law, with conductance κ. If one does not include friction, one can solve the model analytically for an arbitrary working fluid: The temperature of the working fluid should be kept proportional to $(f + \kappa T_{res})^{1/2}$ to maximize the output power. Including friction in the model forces one to resort to numerical solutions for specific working fluids and specific heating functions, $f(t)$.

The optimal path has to balance three factors:

▶ To reduce loss of energy to the heat bath, the working fluid should be cold

▶ To reduce the amount of entropy carried by $f(t)$, the working fluid should be hot

▶ To reduce losses due to friction, the velocity of the piston should be small.

If the pumping heat supply f is a given periodic function of time and if friction is negligible, then the maximum average power delivered by the model is equal to the variance of $(f + \kappa T_{res})^{1/2}$. One can further show that the form of $f(t)$ giving maximum power is a delta function, dumping all the heat into the working fluid at an instant when its temperature is high, again to minimize the concomitant addition of entropy.

Together with several colleagues, we have studied[19] exothermic chemical reactions supplying heat to a mechanical heat engine, using the methods of finite-time thermodynamics. It was no surprise that the reactions convert heat most efficiently when the conversion is infinitely slow and the power output is zero. On the other hand, maximizing the power implies a rate of operation that puts the region of most rapid reaction, the "combustion zone," at the downstream end of the reactor, as anyone who has observed the glow of gases from a Bunsen burner can confirm. The "combustion zone" becomes narrower as the activation energy of the driving chemical reaction increases.

* * *

The research described here that was carried out at the University of Chicago was initially supported by a grant from the National Science Foundation and then by grants from the Department of Energy and the Exxon Educational Foundation. Many of the ideas were generated at workshops conducted at The Aspen Center for Physics, for whose hospitality the authors are indebted.

References

1. M. Mozurkewich, R. S. Berry, Proc. Nat. Acad. Sci. **78**, 1986 (1981); J. Appl. Phys. **53**, 34 (1982).
2. J. Wheatley, T. Hofler, G. W. Swift, A. Migliori, Phys. Rev. Lett. **50**, 499 (1983).
3. For a more theoretical review of many of the topics covered in this article, see B. Andresen, R. S. Berry, M. J. Ondrechen, P. Salamon, "Thermodynamics for Processes in Finite Time," Acc. Chem. Res., in press.
4. P. Salamon, A. Nitzan, B. Andresen, R. S. Berry, Phys. Rev. A **21**, 2115 (1980).
5. P. Salamon, B. Andresen, R. S. Berry, Phys. Rev. A **15**, 2094 (1977).
6. J. Geusic, E. O. Schulz-DuBois, H. E. D. Scovil, Phys. Rev. **156**, 343 (1967).
7. F. Schlögl, Z. Phys. **191**, 81 (1966); Z. Phys. **198**, 559 (1967). I. Procaccia, R. D. Levine, J. Chem. Phys. **65**, 3357 (1976).
8. B. Andresen, P. Salamon, R. S. Berry, J. Chem. Phys. **66**, 1571 (1977).
9. F. L. Curzon, B. Ahlborn, Am. J. Phys. **43**, 22 (1975).
10. P. Salamon, A. Nitzan, J. Chem. Phys. **74**, 3546 (1981).
11. P. Salamon, Y. B. Band, O. Kafri, J. Appl. Phys. **53**, 197 (1982).
12. B. Andresen, M. H. Rubin, R. S. Berry, J. Phys. Chem. **82**, 2704 (1983).
13. P. Salamon, R. S. Berry, Phys. Rev. Lett. **51**, 1127 (1983).
14. F. Weinhold, J. Chem. Phys. **63**, 2479 (1975); J. Chem. Phys. **65**, 559 (1976).
15. M. H. Rubin, Phys. Rev. A **19**, 1272, 1277 (1979).
16. M. H. Rubin, Phys. Rev. A **22**, 1741 (1980).
17. M. J. Ondrechen, B. Andresen, M. Mozurkewich, R. S. Berry, Am. J. Phys. **49**, 681 (1981).
18. Y. B. Band, O. Kafri, and P. Salamon, J. Appl. Phys. **52**, 3745 (1981).
19. M. J. Ondrechen, R. S. Berry, B. Andresen, J. Chem. Phys. **72**, 5118 (1980); M. J. Ondrechen, B. Andresen, R. S. Berry, J. Chem. Phys. **73**, 5838 (1980). □