CONSTRUCTAL THEORY: FROM ENGINEERING DESIGN TO PREDICTING SHAPE AND STRUCTURE IN NATURE

Adrian Bejan
Department of Mechanical Engineering & Materials Science
Duke University,
Durhem, NC, 27708-0300, USA

ABSTRACT

This is a review of two new and important developments in thermal science. First, there exist fundamental optima in the constitution and operation of flow (nonequilibrium) systems, man-made and natural. These optima can be identified based on the simplest models that still retain the essential features of the real systems. Examples are the spatial allocation of heat transfer area in a power plant, and the temporal optimization of on & off processes. The second development is that the engineering method of modeling and optimization has been extended to natural systems, animate and inanimate (e.g., tree networks). This step has been named *constructal theory* for the reasons given in Section 3. The objective of such work is to predict the macroscopic spatial and temporal structure (organization) that is everywhere. It is to inject a dose of determinsm (theory) in a field that until recently considered natural structures to be nondeterministic: results of chance and necessity. These developments bring to mind the advice left to us by J. W. Gibbs more than one hundred years ago:
"One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity."

Key words: Constructal, Fractal, Self-Organization, Non-Equilibrium Thermodynamics.

ENGINEERING ORIGINS

Exciting developments are happening again in thermodynamics. Conceived originally as a theory of heat engine performance, thermodynamics is now called upon to explain the origin and performance of the ultimate in engineering design: Nature itself.

The use of engineering ideas to predict natural organization is the latest development in a revolution that has swept thermodynamics since the energy crisis of the mid-1970s. In the past two decades the theory has been expanded to cover systematically the operation and optimization of real, highly complex systems. The main issue—the driving force in these developments—continues to be performance improvement, efficiency increase, cost minimization, or, simply, good engineering.

The methods of exergy analysis, entropy generation minimization (EGM for short), and thermoeconomics are the lasting results of this revolution. These three methods have been reviewed in a recent book (Bejan *et al.*, 1996). Central to how thermodynamics can shed light on natural organization is the EGM method (Bejan, 1982, 1996a), which is also known as thermodynamic optimization and, more recently in physics, finite-time or finite-size thermodynamics. The EGM method consists of the simultaneous application of principles of thermodynamics, fluid mechanics, and heat and mass transfer. With these principles the analyst constructs *realistic* models, i.e., models that account for the inherent irreversibility of the processes executed by the system and its components.

In the first step, the analysis of the model produces a very important "structural" result: the entropy generation as a function of the size (dimensions), shapes, materials and other physical parameters of the real device. In the second step, the optimization is performed subject to realistic constraints, which are in fact responsible for the irreversible operation of the system.

Let us take a quick look at a few engineering results, so that we may move on to primordial questions that have been sidestepped: Why "optimize" anything? Why do better (e.g., Why less, more, faster, farther, more cheaply, etc.)? Why be "naturally" selected? Why is geometry (shape, structure, similarity) a characteristic of natural flow systems? What is the basis for the increasing complexity (in time) of natural structures? Is there a single physics principle from which geometric form can be *deduced* without any use of empiricism?

DEFROSTING, BREATHING AND TURBULENCE

In engineering and physics, EGM is best represented by its application to models that are the most basic, i.e., the simplest while still realistic. This activity has generated a class of compact results—fundamental optima—that chart the opportunities for design tradeoffs, which deserve to be identified and pursued in practical research and development. These optima point toward strategies for distributing finite material resources in space, and for executing processes in time.

One such example is the selection of the melting material for storing the work content (exergy) of a stream of hot exhaust that is being dumped into the environment. The intercepted exergy is maximal, and the generated entropy is minimal, when the selected material has a melting point equal to the geometric average of the exhaust and environment temperatures (Bejan, 1996a). Another example is the selection of the time interval for heating a single-phase storage material by using a hot stream, which otherwise is discharged into the ambient. The optimal storage time must be such that the heat capacity of the amount of hot material used matches the heat capacity of the amount of storage material.

These fundamental optima require only a few lines of very simple algebra. Even more stunning in this respect is the optimization of the "rhythm" of intermittent processes in which the irreversibility is due to time-dependent diffusion. To illustrate, consider the growth of an ice layer on a cooled surface (Bejan, 1996a). The ice thickness increases as $t_1^{1/2}$, where t_1 is the duration of the freezing process. The rate of ice production decreases as t_1 increases. Clearly, if the objective is to maximize the production of ice (or refrigeration, or exergy storage), it makes sense to interrupt the freezing process, scrape the surface clean, and restart the freezing process. If t_2 is the time interval reserved for cleaning the surface, the time-averaged rate of ice production is proportional to $t_1^{1/2}/(t_1 + t_2)$. This is an important function, which can be maximized by fine-tuning the on & off freezing process: the optimal t_1 equals t_2.

If we look around we find that similar fine-tuning principles work in a wide variety of circumstances, both man-made and natural. In the optimization of the defrost cycle of a refrigerator the objective is again to minimize the generation of entropy (or power input), and the evaporator surface are "scraped" by melting the frost layer. In laminar shear flow the layer thickness increases as $t_1^{1/2}$, and the scraping ("renewal") motion of time t_2 is effected by the eddy. The transition from laminar flow (diffusion) to turbulent flow (streams, organized flow) can now be predicted by simply setting $t_1 = t_2$, that is by maximizing the rate of transport. This holds in every imaginable flow configuration (Bejan, 1993), including Bénard convection.

Jumping ahead to the realm of animate systems, the same principle can be used to anticipate the existence of unique, finely tuned frequencies for breathing and heart beating, which decrease as the body size increases (Bejan, 1997a). These facts have been known empirically for a long time in biology, where they are called alometric laws. They can now be anticipated theoretically by starting from engineering thermodynamics and EGM. The larger issue is to understand the purpose of the optimization that rules the naturally occurring structure. Why should "structure" occur naturally, and why should it optimize itself? To these questions I return in the second part of this article.

Organization and optimization occur not only in time but also in space. Consider an actual (irreversible) power plant that operates between the high temperature T_H and low temperature T_L, Fig. 1. The plant owes its imperfection to the heat exchangers that facilitate the two heat interactions (Bejan, 1982, 1996a). The rest of the power plant is assumed irreversibility free. Each heat interaction is proportional to the temperature difference and the size of the heat exchanger: $Q_H = C_H \Delta T_H$ and $Q_L = C_L \Delta T_L$. The total heat-exchange inventory is fixed, $C_H + C_L$ = constant. When the heat input Q_H is *fixed*, the power output W is maximal when the hardware is partitioned equally between the two ends of the power plant, $C_H = C_L$. This *principle of equipartition* (spatial allocation, distribution) of resources is a characteristic of many other thermodynamically optimized systems.

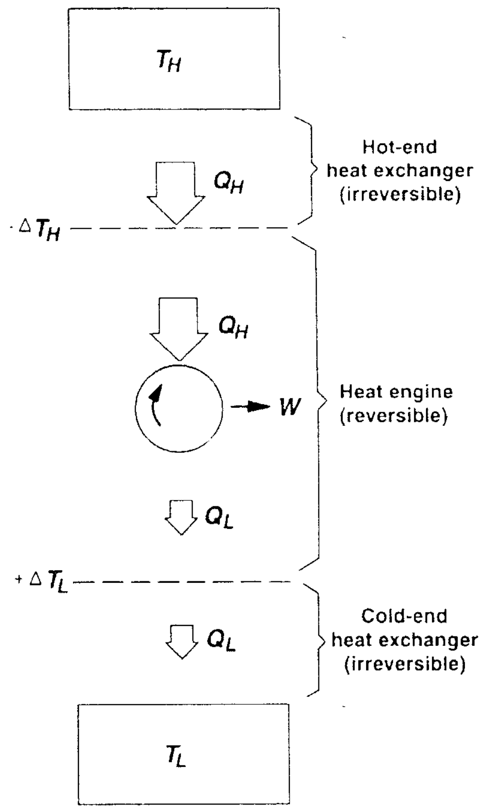

Fig. 1 - Spatial allocation of heat-exchange hardware in a simple power plant model with two temperature differences (Bejan, 1982).

A word of caution is needed here, because the simplicity of models such as Fig. 1 can trick us into using them in inappropriate situations. One example—a famous one—is the use of the same power plant model in combination with the new assumption that the heat input Q_H is free to vary, i.e., as free as the rejected heat Q_L. In this case the design has two degrees of freedom, hence two optima: (1) the allocation of hardware, the optimization of which leads again to equipartition ($C_H = C_L$), and (2) the inner temperature ratio $(T_H - \Delta T_H)/(T_L + \Delta T_L)$, the optimal value of which is $(T_H/T_L)^{1/2}$ for which the efficiency is constant, $1 - (T_L/T_H)^{1/2}$, and considerably lower than the Carnot ceiling, $1 - T_L/T_H$. Gyftopoulos (1997) noted, that result (2) is false for several reasons, starting with the history of heat engine development which is one characterized by efficiencies that continue to increase, in time. This, by the way, is the natural tendency—the same type of empirical observation—that supports the constructal law of macroscopic natural organization, which concludes this article.

TREE NETWORKS IN NATURE

Thermodynamic optimization can deliver not only the optimal distribution of material but also the optimal dimensions of components. For example, in the most elementary passage of a heat exchanger the entropy generation is due to both heat transfer and fluid friction. These two contributions compete against one another. The hydraulic diameter of the passage can be selected such that the sum of the two irreversibilities is minimal. The dimensions of bodies immersed in external convection can be selected similarly (Bejan, 1982).

Even simpler is the sizing of a system that owes its irreversibility to only one transport mechanism, e.g., heat transfer. If the heat current is imposed, the minimization of entropy generation reduces to minimizing the resistance to heat flow. In the cooling of electronic packages the volume is fixed, and so is the heat generation rate that is distributed uniformly over the volume. The geometric arrangement of the heat generating components can be optimized such that the hot-spot temperature is minimal (Bejan, 1995). The spacing or number of components is free to vary. If the spacing is too large, there is not enough heat transfer area and the hot-spot temperature is high. When the spacing is too small, the coolant flow rate decreases, and the hot-spot temperature is again high. There is an optimal spacing—an optimal package architecture—that minimizes the thermal resistance between the overall system and ambient. This geometric principle finds wide applicability in both mad-made systems (e.g., computers) and natural systems (e.g., mud cracks) (Bejan, 1997b).

A powerful geometric principle was discovered recently in the minimization of the thermal resistance between a heat-generating volume and one point (Bejan, 1997c). The volume is fixed, and the heat-generating material has a low thermal conductivity (k_0). A small amount of high-conducting material (k_p) is to be distributed through the k_0 material such that the overall volume-to-point resistance is minimal.

The discovery is that every portion—every subsystem of the given volume—can have its shape optimized. Figure 2 illustrates this principle at the smallest volume scale, where a single k_p fiber removes from the system the heat generated by the k_0 material. The optimal rectangular shape that minimizes the thermal resistance between the element and the exit end of its k_p fiber is $(H_0/L_0)_{opt} = 2(k_0H_0/k_pD_0)^{1/2}$.

The same geometric optimization principle applies at larger scales. The larger volume element is an assembly—a construct—of optimized volume elements of the smallest size, Fig. 3. This construct too can have its shape (or number of constituents) optimized. This single principle of construction and shape optimization continues toward stepwise larger scales until the given volume is covered. The end result is shape and structure—the optimized architecture of the composite (k_0, k_p) that connects the sink point to the finite-size volume (Bejan, 1997c; Ledezma *et al.*, 1997).

The infinity of points of the given volume is "connected" to the sink because at the smallest volume scale the transport is volumetric, by thermal diffusion through the low-conductivity material. At larger scales the transport is via channels (streams) of higher conductivity. Diffusion comes first and streams later.

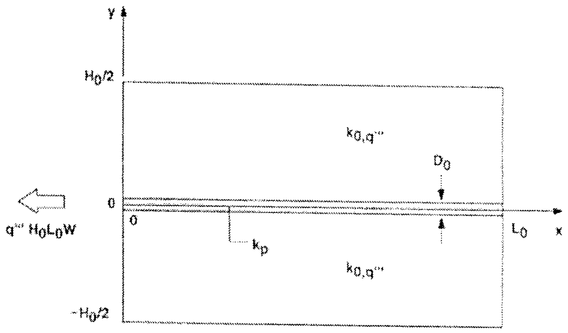

Fig. 2 - Slender elemental volume with volumetric heat generating and one high-conductivity path along its axis of symmetry (Bejan, 1997c).

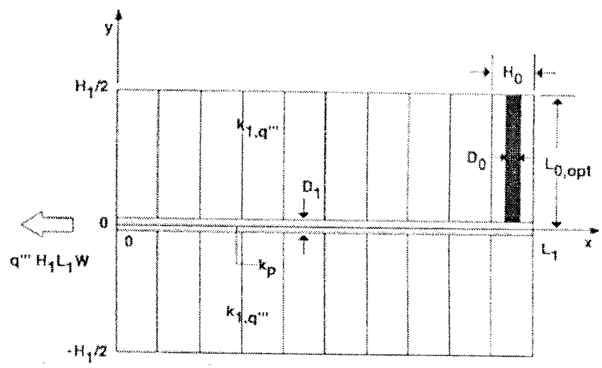

Fig. 3 - The first construct: a large number of elemental volumes (Fig. 2) connected to a central high conductivity path (Bejan, 1997c).

This complete volume-to-point connection is a first in mathematics. A number of points could be connected optimally to a single point, by computer. As computers become more powerful, larger numbers of such points will be connected in the future. Their number, however, will never by infinite *to fill* the given volume. Furthermore, the opaque (black box) optimization executed by the computer will never be theory.

Astonishing is not that the high-conductivity channels form a tree (a network without loops) but that each feature of the tree is deterministic, the result of a single principle of optimization. This conclusion runs against the currently accepted doctrine that natural structures are nondeterministic, the result of chance and necessity. Fractal geometry too is descriptive, not predictive. Any tree can be simulated by repeating an assumed algorithm, and truncating this operation at an arbitrary, small (finite-size) scale.

The discovery then is not the tree but the constructal principle that generates this structure, from one scale to the next larger scale. The *name constructal** emphasizes the link between determinism (theory) and the direction from small to large. It is also a reminder that theory runs counter to fractal thinking.

Tree networks abound in Nature, in both animate and inanimate flow systems. We find them everywhere: botanical trees, leaves, roots, lungs, vascularized tissues, neural dendrites, river drainage basins, river deltas, lightning and dendritic crystals (*dendron* means tree in Greek). Every detail of every natural tree can be anticipated trough the construction and optimization shown for the "heat tree" in Figs. 2 and 3. In fluid trees the small scale volumetric flow is by slow viscous diffusion (e.g., Darcy flow in the wet banks of the smallest rivulet), while the larger-scale flow is organized into faster conduits: streams (Bejan, 1997b,d).

THE CONSTRUCTAL LAW OF SHAPE AND STRUCTURE IN NATURE

The volume-to-point constructs have a definite time direction: from small to large, and from shapelessness (diffusion) to structure (channels, streams). Determinism results only if this time arrow is respected. If the reversed time direction is used, i.e., from large to small, through the repeated fracturing of a postulated network into smaller and smaller pieces (as in fractal geometry), then it is impossible to predict theoretically the optimal volume-to-point flow architecture (Bejan, 1996b, 1997c).

The optimized geometry formed by the slow and fast flow regimes unites all the volume-to-point flows. Think of the flow of oxygen through a mammal: the slow, shapeless flow is volumetric mass diffusion through the tissues, while the faster regime is mass convection (streams) through blood vessels and bronchiai passages. Think also of turbulent flow: diffusion in the smallest volume elements is accompanied by the structure of faster streams known as eddies. Artificial constructs such as the internal arrangement of computers require the same cooperation between slow and fast heat transfer, with the slow mode placed at the smallest scale. This cooperation is most obvious in living groups, from bacterial colonies to urban growth (Bejan, 1996b): every member has a place in the structure, in such a way that every member benefits. The urge to organize is an expression of selfish behavior.

The constructal nature of access optimization is even more obvious in the context of minimizing the time of travel between one point and a finite-size area (an infinity of points) (Bejan, 1996b). Travelers have access to more than one mode of locomotion, starting with the slowest speed (V_0, walking) and proceeding toward faster modes (vehicles, $V_1 < V_2 < ...$). The given area is covered in steps of increasingly larger constructs (A_1, A_2, A_3,...). Each construct can be optimized for overall shape and angle between assembly and constituents. For example, Fig. 4 shows the smallest area element (A1), for which the optimal shape is $(H_1/L_1)opt = 2V_0/V_1$, and where the angle between V_0 and V_1 was set at 90° (a good approximation when $V_0 << V_1$).

The shape-optimization process repeats itself at larger area scales (Fig. 5). Urban growth is predicted, from backyards and alleys, to streets, avenues and highways. Along the way, we also discover the long sought *principle*, i.e., the reason why natural tree networks "happen to look" like the truncated fractal images postulated by the mathematician. Constructal theory is about the physics that had been missing in fractals (Kadanoff, 1986).

The minimization of travel time has been invoked in the past to account for the shape of light rays. When the ray strikes a mirror, the optimal angle of incidence equals the angle of reflection (Heron of Alexandria). Furthermore, the ray is bent to an optimal angle as it passes from one medium into another (Fermat). The time or resistance minimization principle is raised to the rank of law by the structure of light rays and all the tree networks anticipated by constructal theory. This law can be summarized as follows (Bejan, 1997b,c).

"For a finite-size system to persist in time (to live), it must evolve in such a way that it provides easier access to the imposed currents that flow through it."

This statement has two parts. First, it recognizes the natural tendency of imposed currents to construct paths of optimal access (e.g., shapes, structures) through constrained open systems. The second part accounts for the evolution (i.e., improvements) of these paths, which occurs in an identifiable direction that can be aligned with time itself.

This constructal principle accounts for the choices that are made by macroscopic open systems (natural or man-made) subjected to flow and size constraints. The cross-section of the blood vessel is nearly round for the same reason that the width of the river is proportional to the depth (Bejan, 1997b). This law is about defining the very concept of "optimization", or "purpose". It bridges the gap not only between physics and biology with its many related fields such as economics, but also the gap between physics and engineering.

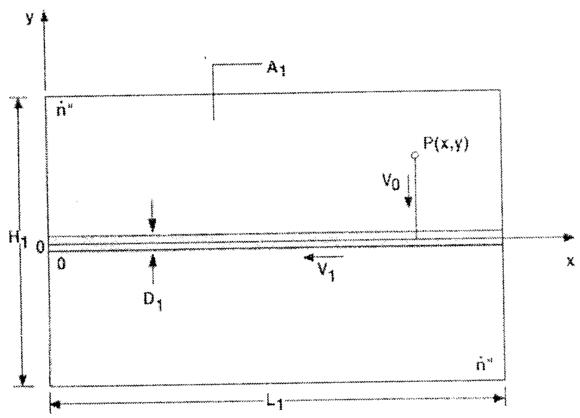

Fig. 4 - The smallest (innermost) elemental area, A_1, and the street segment allocated to it (Bejan, 1996b).

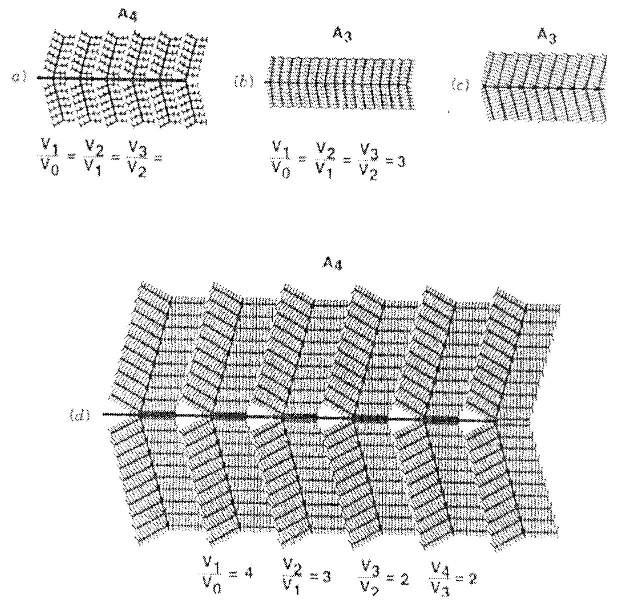

Fig. 5 - The growth of street patterns, as the minimization of travel time between a finite-size area and one point (Bejan & Ledezma, 1998). Velocities increase as the constructs become larger: $V_0 < V_1 < V_2 < V_3 < V_4$. In these three examples, each construct has been optimized for both shape and angle.

Constructal theory introduces an engineering flavor into the current debate on natural order, which until now has been carried out mainly in physics and biology. As a result of their training, engineers begin the design of a device by first understanding its purpose. The size of the device is always finite, never infinitesimal. The device must function (i.e., fulfill its purpose) subject to certain constraints. To analyze the device is not sufficient: to optimize it, to construct it, and to make it work is the real objective. Finally, many designs that differ in some of the finer details have nearly the same overall performance as the optimal design (Bejan et al., 1996).

All these featurespurpose, finite size, constraints, optimization, constructioncan be seen in the animate and inanimate structures that surround us. It is time that we engineers expand the determinist powers of our thermodynamics over the field of naturally organized systems. We are the ones to do this work, because Nature is engineered, not random. My own progress in this direction is described in a new book (Bejan, 1997b).

Acknowledgment. This research was supported by the National Science Foundation.

REFERENCES

Bejan, A. 1982. *Entropy Generation through Heat and Fluid Flow*, Wiley, New York.

Bejan, A. 1993. *Heat Transfer*, Wiley, New York.

Bejan, A. 1995. *Convection Heat Transfer, Second Edition*, Wiley, New York.

Bejan, A. 1996a. Entropy Generation Minimization, CRC Press, Boca Raton, FL.

Bejan, A. 1996b. Street Network Theory of Organization in Nature, *Journal of Advanced Transportation*, v. 30(2), p. 85-107.

Bejan, A. 1997a. *Theory of Organization in Nature:* Pulsating Physiological Processes, International Journal of Heat and Mass Transfer, vol. 40, pp. 2097-2104.

Bejan A. 1997b. *Advanced Engineering Thermodynamics*, Second Edition, Wiley, New York.

Bejan, A. 1997c. *Constructal-Theory Network of Conducting Paths for Cooling a Heat Generating Volume,* International Journal of Heat and Mass Transfer, v. 40, p. 799-816.

Bejan, A. 1997d. *Constructal Tree Network for Fluid Flow Between a Finite-Size Volume and One Source of Sink*, Revue Générale de Thermique, v. 36(8), p. 592-604.

Bejan, A. and Ledezma, G. A. 1998. *Streets Tree Networks and Urban Growth*: Optimal Geometry for Quickest Access Between a Finite-Size Volume and One Point, Physica A, v. 255, pp. 211-217.

Bejan, A. Tsatsaronis, G. and Moran, M. 1996. *Thermal Design and Optimization, Wiley*, New York.

Gyftopoulos. E. P. 1997. *Fundamentals of Analyses and Processes, Energy Conversion and Management*, v. 38(15-17), p. 1525-1533.

Kadanoff, L. P. 1986. Fractals: *Where's the Physics?* Physics Today, pp. 6-7, February.

Ledezma, G.A. Bejan, A. and Errera, M. R. 1997. *Constructal Tree Networks for Heat Transfer*, Journal of Applied Physics, v. 82(1), p. 89-100.